57410

STIC-Biotech/ChemLib

From: Schnizer, Holly
Sent: Wednesday, January 02, 2002 1:21 PM
To: STIC-Biotech/ChemLib
Subject: seq. search request for appl. no. 09/444,281

Please search the commercial and interference databases for

SEQ ID NOs: 27, 35, and 36 and polynucleotides encoding SEQ ID NOs: 27, 35, and 36

Thank you.

Holly Schnizer
AU 1653
CM1-10B05
305-3722
mailbox: CM1-9B01

CRF

Point of Contact:
Searcher: Sheppard
Phone:
Location: Tel: 308-4499
Date Picked Up:
Date Completed: 1/8/02
Searcher Prep/Review:
Clerical:
Online time:

TYPE OF SEARCH:
NA Sequences:
AA Sequences:
Structures:
Bibliographic:
Litigation:
Full text:
Patent Family:
Other:

VENDOR/COST(where applic.)
STN:
DIALOG:
Questel/Orbit:
DRLink:
Lexis/Nexis:
Sequence Sys.:
WWW/Internet:
Other (specify):

PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT toxicity, useful therapeutically, as disinfectants etc.
XX
PS Claim 11; Page 88; 129pp; English.
XX
CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC (II), BBBXZXXZXB (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC (V), LBBnXZnXXZnXRK (VI), LKnXZXXZXRRK (VII) and BBXZXXZXBBB (VIII).
CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC infections caused by bacteria (Gram positive or negative, or anaerobic);
CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC trematodes) or viruses. Typical of very many pathogens that can be
CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC derived from the analogues may be used similarly; the compounds may
CC also be prepared from antibiotics or antiarrythmic agents. The analogues
CC may be used therapeutically or to coat medical devices; also they are
CC useful as surface disinfectants, as additives to shampoo or soaps, as
CC insecticides or herbicides, or as preservatives for foods and technical
CC materials. The analogues are administered by injection, lavage, orally
CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC spectrum of activity than indolicidin and modification as compounds
CC reduces their toxicity.
XX
SQ Sequence 12 AA;

Query Match 100.0%; Score 86; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 2
AAY94496
ID AAY94496 standard; Peptide; 12 AA.
XX
AC AAY94496;
XX
DT 20-SEP-2000 (first entry)
XX
DE MBI-11B7 peptide derived from indolicidin.
XX
KW Cellulose binding domain; CBD; cationic peptide;
KW MBI-11B7; indolicidin; bovine.
XX
OS Bos taurus.
XX
PN WO200031279-A2.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-CA01107.
XX
PR 20-NOV-1998; 98US-0109218.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Burian J, Bartfeld D;
XX
DR WPI; 2000-400086/34.
XX
PT Multi-domain fusion protein expression cassette used for high yield
PT stable production of foreign peptide gene products -
XX
PS Disclosure; Page 24; 73pp; English.
XX
CC A novel method allows the efficient production of cationic peptides in
CC recombinant host cells. The method involves construction of a
CC multi-domain fusion protein expression cassette comprising a promoter and
CC a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC of anionic peptide sequences in the linker sequences neutralises the
CC positive charge of the cationic peptide so that the charge of the
CC fusion protein is controlled. This cassette allows high yield, stable
CC production of the cationic peptide. Cationic peptides such as
CC bovine indolicidin may be used as antimicrobial agents. The present
CC sequence is the MBI-11B7 peptide. MBI-11B7 is a cationic peptide derived
CC from modifications of indolicidin.
XX
SQ Sequence 12 AA;

Query Match 100.0%; Score 86; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 3
AAY91791
ID AAY91791 standard; Peptide; 12 AA.
XX
AC AAY91791;
XX
DT 06-JUN-2000 (first entry)
XX
DE Amino acid sequence of cationic peptide MBI 11B7CN.
XX
KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW multidrug resistance.
XX
OS Synthetic.
XX
PN WO9965506-A2.
XX
PD 23-DEC-1999.
XX
PF 14-JUN-1999; 99WO-CA00552.
XX
PR 12-JUN-1998; 98US-0096541.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Claim 1; Page 14; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX
SQ Sequence 12 AA;

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:40:26 ; Search time 53.46 Seconds
(without alignments)
16.627 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRWPWWPWRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	100.0	12	19	AAY24550	Indolicidin analog
2	86	100.0	12	21	AAY94496	MBI-11B7 peptide d
3	86	100.0	12	21	AAY91791	Amino acid sequenc
4	86	100.0	20	19	AAY24553	Indolicidin analog
5	86	100.0	20	21	AAY91797	Amino acid sequenc
6	86	100.0	21	19	AAY24552	Indolicidin analog
7	86	100.0	21	19	AAY24554	Indolicidin analog
8	86	100.0	21	19	AAW66376	Cationic peptide o
9	86	100.0	21	21	AAY91796	Amino acid sequenc
10	86	100.0	21	21	AAY91798	Amino acid sequenc
11	86	100.0	27	19	AAW66363	Indolicidin analog
12	86	100.0	28	21	AAY91800	Amino acid sequenc
13	83	96.5	12	19	AAY24567	Indolicidin analog
14	83	96.5	12	21	AAY91788	Amino acid sequenc
15	82	95.3	12	19	AAY24594	Indolicidin analog
16	82	95.3	12	19	AAW66364	Indolicidin analog
17	82	95.3	12	21	AAY91817	Amino acid sequenc
18	82	95.3	12	21	AAY91841	Amino acid sequenc
19	81	94.2	12	19	AAY24605	Indolicidin analog
20	81	94.2	12	19	AAY24595	Indolicidin analog
21	81	94.2	12	21	AAY91842	Amino acid sequenc
22	81	94.2	12	21	AAY91852	Amino acid sequenc
23	80	93.0	12	19	AAY24596	Indolicidin analog
24	80	93.0	12	19	AAY24603	Indolicidin analog
25	80	93.0	12	19	AAY24604	Indolicidin analog
26	80	93.0	12	21	AAY91843	Amino acid sequenc
27	80	93.0	12	21	AAY91850	Amino acid sequenc
28	80	93.0	12	21	AAY91851	Amino acid sequenc
29	78	90.7	12	19	AAY24598	Indolicidin analog
30	78	90.7	12	19	AAY24601	Indolicidin analog
31	78	90.7	12	19	AAW66361	Indolicidin analog
32	78	90.7	12	21	AAY91785	Amino acid sequenc
33	78	90.7	12	21	AAY91845	Amino acid sequenc
34	78	90.7	12	21	AAY91848	Amino acid sequenc
35	78	90.7	13	19	AAY24565	Indolicidin analog
36	78	90.7	13	21	AAY91786	Amino acid sequenc
37	77	89.5	12	19	AAY24586	Indolicidin analog
38	77	89.5	12	21	AAY91828	Amino acid sequenc
39	75	87.2	11	19	AAY24569	Indolicidin analog
40	75	87.2	11	21	AAY91790	Amino acid sequenc
41	75	87.2	12	19	AAY24580	Indolicidin analog
42	75	87.2	12	21	AAY91804	Amino acid sequenc
43	75	87.2	13	18	AAW12873	Antimicrobial cati
44	75	87.2	13	18	AAW12895	Antimicrobial cati
45	75	87.2	13	18	AAW12896	Antimicrobial cati

ALIGNMENTS

RESULT 1
AAY24550
ID AAY24550 standard; peptide; 12 AA.
XX
AC AAY24550;
XX
DT 18-AUG-1999 (first entry)
XX
DE Indolicidin analogue #2.
XX
KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW additive; shampoo; soap; insecticide; herbicide; preservative;
KW food; technical material.
XX
OS Synthetic.
XX
PN WO9807745-A2.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-US14779.
XX
PR 13-JAN-1997; 97US-0034949.
PR 21-AUG-1996; 96US-0024754.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Erfle D, Fraser JR, Krieger TJ, Taylor R, West MH;
XX
DR WPI; 1998-169090/15.
XX
PT New indolicidin analogues with antimicrobial activity and related
PT nucleic acid - vectors, transformed cells and antibodies, also

Query Match 100.0%; Score 86; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 4
AAY24553
ID AAY24553 standard; peptide; 20 AA.
XX
AC AAY24553;
XX
DT 18-AUG-1999 (first entry)
XX
DE Indolicidin analogue #5.
XX
KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW additive; shampoo; soap; insecticide; herbicide; preservative;
KW food; technical material.
XX
OS Synthetic.
XX
PN WO9807745-A2.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-US14779.
XX
PR 13-JAN-1997; 97US-0034949.
PR 21-AUG-1996; 96US-0024754.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Erfle D, Fraser JR, Krieger TJ, Taylor R, West MH;
XX
DR WPI; 1998-169090/15.
XX
PT New indolicidin analogues with antimicrobial activity and related
PT nucleic acid - vectors, transformed cells and antibodies, also
PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT toxicity, useful therapeutically, as disinfectants etc.
XX
PS Claim 11; Page 88; 129pp; English.
XX
CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC (II), BBBXZXXZXB (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC (V), LBBnXZnXXZnXRK (VI), LKnXZXXZXRRK (VII) and BBXZXXZXBBB (VIII).
CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC infections caused by bacteria (Gram positive or negative, or anaerobic);
CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC trematodes) or viruses. Typical of very many pathogens that can be
CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC derived from the analogues may be used similarly; the compounds may
CC also be prepared from antibiotics or antiarrythmic agents. The analogues
CC may be used therapeutically or to coat medical devices; also they are
CC useful as surface disinfectants, as additives to shampoo or soaps, as
CC insecticides or herbicides, or as preservatives for foods and technical
CC materials. The analogues are administered by injection, lavage, orally
CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC spectrum of activity than indolicidin and modification as compounds
CC reduces their toxicity.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 86; DB 19; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 5
AAY91797
ID AAY91797 standard; Peptide; 20 AA.
XX
AC AAY91797;
XX
DT 06-JUN-2000 (first entry)
XX
DE Amino acid sequence of cationic peptide MBI 11B17CN.
XX
KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW multidrug resistance.
XX
OS Synthetic.
XX
PN WO9965506-A2.
XX
PD 23-DEC-1999.
XX
PF 14-JUN-1999; 99WO-CA00552.
XX
PR 12-JUN-1998; 98US-0096541.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Disclosure; Page 15; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 86; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 6
AAY24552

```
ID   AAY24552 standard; peptide; 21 AA.
XX
AC   AAY24552;
XX
DT   18-AUG-1999  (first entry)
XX
DE   Indolicidin analogue #4.
XX
KW   Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW   antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW   additive; shampoo; soap; insecticide; herbicide; preservative;
KW   food; technical material.
XX
OS   Synthetic.
XX
PN   WO9807745-A2.
XX
PD   26-FEB-1998.
XX
PF   21-AUG-1997;   97WO-US14779.
XX
PR   13-JAN-1997;   97US-0034949.
PR   21-AUG-1996;   96US-0024754.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Erfle D,  Fraser JR,  Krieger TJ,  Taylor R,  West MH;
XX
DR   WPI; 1998-169090/15.
XX
PT   New indolicidin analogues with antimicrobial activity and related
PT   nucleic acid - vectors, transformed cells and antibodies, also
PT   conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT   toxicity, useful therapeutically, as disinfectants etc.
XX
PS   Claim 11; Page 88; 129pp; English.
XX
CC   AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC   (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC   (II), BBBXZXXZXB  (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC   (V), LBBnXZnXXZnXRK  (VI), LKnXZXXZXRRK  (VII) and BBXZXXZXBBB  (VIII).
CC   Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC   preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC   in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC   infections caused by bacteria (Gram positive or negative, or anaerobic);
CC   fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC   trematodes) or viruses. Typical of very many pathogens that can be
CC   controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC   hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC   aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC   derived from the analogues may be used similarly; the compounds may
CC   also be prepared from antibiotics or antiarrythmic agents. The analogues
CC   may be used therapeutically or to coat medical devices; also they are
CC   useful as surface disinfectants, as additives to shampoo or soaps, as
CC   insecticides or herbicides, or as preservatives for foods and technical
CC   materials. The analogues are administered by injection, lavage, orally
CC   or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC   spectrum of activity than indolicidin and modification as compounds
CC   reduces their toxicity.
XX
SQ   Sequence   21 AA;

  Query Match             100.0%;  Score 86;  DB 19;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||||||||||||
Db       1 ilrwpwwpwrrk 12


RESULT   7
```

```
AAY24554
ID   AAY24554 standard; peptide; 21 AA.
XX
AC   AAY24554;
XX
DT   18-AUG-1999  (first entry)
XX
DE   Indolicidin analogue #6.
XX
KW   Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW   antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW   additive; shampoo; soap; insecticide; herbicide; preservative;
KW   food; technical material.
XX
OS   Synthetic.
XX
PN   WO9807745-A2.
XX
PD   26-FEB-1998.
XX
PF   21-AUG-1997;   97WO-US14779.
XX
PR   13-JAN-1997;   97US-0034949.
PR   21-AUG-1996;   96US-0024754.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Erfle D,  Fraser JR,  Krieger TJ,  Taylor R,  West MH;
XX
DR   WPI; 1998-169090/15.
XX
PT   New indolicidin analogues with antimicrobial activity and related
PT   nucleic acid - vectors, transformed cells and antibodies, also
PT   conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT   toxicity, useful therapeutically, as disinfectants etc.
XX
PS   Claim 11; Page 88; 129pp; English.
XX
CC   AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC   (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC   (II), BBBXZXXZXB  (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC   (V), LBBnXZnXXZnXRK  (VI), LKnXZXXZXRRK  (VII) and BBXZXXZXBBB  (VIII).
CC   Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC   preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC   in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC   infections caused by bacteria (Gram positive or negative, or anaerobic);
CC   fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC   trematodes) or viruses. Typical of very many pathogens that can be
CC   controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC   hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC   aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC   derived from the analogues may be used similarly; the compounds may
CC   also be prepared from antibiotics or antiarrythmic agents. The analogues
CC   may be used therapeutically or to coat medical devices; also they are
CC   useful as surface disinfectants, as additives to shampoo or soaps, as
CC   insecticides or herbicides, or as preservatives for foods and technical
CC   materials. The analogues are administered by injection, lavage, orally
CC   or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC   spectrum of activity than indolicidin and modification as compounds
CC   reduces their toxicity.
XX
SQ   Sequence   21 AA;

  Query Match             100.0%;  Score 86;  DB 19;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||||||||||||
Db       1 ilrwpwwpwrrk 12
```

```
RESULT   8
AAW66376
ID   AAW66376 standard; peptide; 21 AA.
XX
AC   AAW66376;
XX
DT   12-JAN-1999  (first entry)
XX
DE   Cationic peptide of claim 15 #3.
XX
KW   Indolicidin analogue; resistance; cationic peptide; antibiotic;
KW   bacterial infection; tolerance; antibacterial; microorganism;
KW   bacteria; fungus; parasite; virus.
XX
OS   Synthetic.
XX
PN   WO9840401-A2.
XX
PD   17-SEP-1998.
XX
PF   10-MAR-1998;   98WO-CA00190.
XX
PR   25-FEB-1998;   98US-0030619.
PR   10-MAR-1997;   97US-0040649.
PR   20-AUG-1997;   97US-0915314.
PR   26-SEP-1997;   97US-0060099.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Fraser JR,  McNicol PJ,  West MHP;
XX
DR   WPI; 1998-520800/44.
XX
PT   New indolicidin peptide analogues - useful for, e.g. enhancing
PT   activity of antibiotic or overcoming tolerance, acquired resistance
PT   or inherent resistance of microorganisms
XX
PS   Claim 15; Page 93; 105pp; English.
XX
CC   The present sequence represents a specifically claimed cationic peptide
CC   from the present invention. The present invention describes compositions
CC   and methods for treating infection, especially bacterial infections. The
CC   compositions and methods use cationic peptides in combination with an
CC   antibiotic agent which are then administered to a patient to enhance the
CC   activity of the antibiotic agent, to overcome: (a) tolerance; (b)
CC   acquired resistance; and (c) inherent resistance. The combinations of
CC   antibiotics and cationic peptides can provide synergistic activity
CC   against a microorganism that is tolerant, inherently resistant, or has
CC   acquired resistance to an antibiotic agent. They can be used for killing
CC   e.g. bacteria, fungi, parasites and viruses.
XX
SQ   Sequence   21 AA;


  Query Match             100.0%;  Score 86;  DB 19;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILRWPWWPWRRK 12
          ||||||||||||
Db      1 ilrwpwwpwrrk 12


RESULT   9
AAY91796
ID   AAY91796 standard; Peptide; 21 AA.
XX
AC   AAY91796;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11B16CN.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Friedland HD,  Krieger TJ,  Taylor R,  Erfle D,  Fraser JR,  West MHP;
XX
DR   WPI; 2000-223549/19.
XX
PT   Novel pharmaceutical composition containing optionally activated
PT   polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT   -
XX
PS   Disclosure; Page 15; 94pp; English.
XX
CC   This sequence represents a cationic peptide amino acid sequence, which
CC   can be used in the pharmaceutical composition of the invention. The
CC   invention relates to a pharmaceutical composition containing at least one
CC   activated polyoxyalkylene (APO)-modified cationic peptide. The
CC   modification of peptides with APO increases their activity against tumour
CC   cells, including those with a multidrug resistant phenotype. The
CC   pharmaceutical composition can be used to treat tumours, specifically
CC   lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC   cervix, uterus, skin, prostate, liver and colon.
XX
SQ   Sequence   21 AA;


  Query Match             100.0%;  Score 86;  DB 21;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILRWPWWPWRRK 12
          ||||||||||||
Db      1 ilrwpwwpwrrk 12


RESULT   10
AAY91798
ID   AAY91798 standard; Peptide; 21 AA.
XX
AC   AAY91798;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11B18CN.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
```

XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Disclosure; Page 15; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 86; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.8e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 11
AAW66363
ID AAW66363 standard; peptide; 27 AA.
XX
AC AAW66363;
XX
DT 12-JAN-1999 (first entry)
XX
DE Indolicidin analogue MBI 11B20.
XX
KW Indolicidin analogue; resistance; cationic peptide; antibiotic;
KW bacterial infection; tolerance; antibacterial; microorganism;
KW bacteria; fungus; parasite; virus.
XX
OS Bos taurus.
OS Synthetic.
XX
PN WO9840401-A2.
XX
PD 17-SEP-1998.
XX
PF 10-MAR-1998; 98WO-CA00190.
XX
PR 25-FEB-1998; 98US-0030619.
PR 10-MAR-1997; 97US-0040649.
PR 20-AUG-1997; 97US-0915314.
PR 26-SEP-1997; 97US-0060099.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Fraser JR, McNicol PJ, West MHP;
XX
DR WPI; 1998-520800/44.
XX
PT New indolicidin peptide analogues - useful for, e.g. enhancing
PT activity of antibiotic or overcoming tolerance, acquired resistance
PT or inherent resistance of microorganisms
XX
PS Claim 1; Page 91; 105pp; English.
XX
CC The present sequence represents an indolicidin analogue. The present
CC invention describes compositions and methods for treating infection,
CC especially bacterial infections. The compositions and methods use
CC cationic peptides in combination with an antibiotic agent which are
CC then administered to a patient to enhance the activity of the antibiotic
CC agent, to overcome: (a) tolerance; (b) acquired resistance; and (c)
CC inherent resistance. The combinations of antibiotics and cationic
CC peptides can provide synergistic activity against a microorganism that
CC is tolerant, inherently resistant, or has acquired resistance to an
CC antibiotic agent. They can be used for killing e.g. bacteria, fungi,
CC parasites and viruses.
XX
SQ Sequence 27 AA;

Query Match 100.0%; Score 86; DB 19; Length 27;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRRK 12
||||||||||||
Db 1 ilrwpwwpwrrk 12

RESULT 12
AAY91800
ID AAY91800 standard; Peptide; 28 AA.
XX
AC AAY91800;
XX
DT 06-JUN-2000 (first entry)
XX
DE Amino acid sequence of cationic peptide MBI 11B20CN.
XX
KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW multidrug resistance.
XX
OS Synthetic.
XX
PN WO9965506-A2.
XX
PD 23-DEC-1999.
XX
PF 14-JUN-1999; 99WO-CA00552.
XX
PR 12-JUN-1998; 98US-0096541.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Claim 1; Page 15; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX

```
SQ   Sequence   28 AA;

  Query Match           100.0%;  Score 86;  DB 21;  Length 28;
  Best Local Similarity  100.0%;  Pred. No. 9.1e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||||||||||||
Db       1 ilrwpwwpwrrk 12


RESULT  13
AAY24567
ID   AAY24567 standard; peptide; 12 AA.
XX
AC   AAY24567;
XX
DT   18-AUG-1999  (first entry)
XX
DE   Indolicidin analogue #19.
XX
KW   Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW   antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW   additive; shampoo; soap; insecticide; herbicide; preservative;
KW   food; technical material.
XX
OS   Synthetic.
XX
PN   WO9807745-A2.
XX
PD   26-FEB-1998.
XX
PF   21-AUG-1997;   97WO-US14779.
XX
PR   13-JAN-1997;   97US-0034949.
PR   21-AUG-1996;   96US-0024754.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Erfle D,  Fraser JR,  Krieger TJ,  Taylor R,  West MH;
XX
DR   WPI; 1998-169090/15.
XX
PT   New indolicidin analogues with antimicrobial activity and related
PT   nucleic acid - vectors, transformed cells and antibodies, also
PT   conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT   toxicity, useful therapeutically, as disinfectants etc.
XX
PS   Claim 12; Page 89; 129pp; English.
XX
CC   AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC   (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC   (II), BBBXZXXZXB  (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC   (V), LBBnXZnXXZnXRRK  (VI), LKnXZXXZXRRK  (VII) and BBXZXXZXBBB  (VIII).
CC   Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC   preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC   in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC   infections caused by bacteria (Gram positive or negative, or anaerobic);
CC   fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC   trematodes) or viruses. Typical of very many pathogens that can be
CC   controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC   hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC   aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC   derived from the analogues may be used similarly; the compounds may
CC   also be prepared from antibiotics or antiarrythmic agents. The analogues
CC   may be used therapeutically or to coat medical devices; also they are
CC   useful as surface disinfectants, as additives to shampoo or soaps, as
CC   insecticides or herbicides, or as preservatives for foods and technical
CC   materials. The analogues are administered by injection, lavage, orally
CC   or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC   spectrum of activity than indolicidin and modification as compounds
CC   reduces their toxicity.
XX
SQ   Sequence   12 AA;

  Query Match            96.5%;  Score 83;  DB 19;  Length 12;
  Best Local Similarity  91.7%;  Pred. No. 1e-05;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||:|||||||||
Db       1 ilkwpwwpwrrk 12


RESULT  14
AAY91788
ID   AAY91788 standard; Peptide; 12 AA.
XX
AC   AAY91788;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11B3CN.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Friedland HD,  Krieger TJ,  Taylor R,  Erfle D,  Fraser JR,  West MHP;
XX
DR   WPI; 2000-223549/19.
XX
PT   Novel pharmaceutical composition containing optionally activated
PT   polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT   -
XX
PS   Disclosure; Page 14; 94pp; English.
XX
CC   This sequence represents a cationic peptide amino acid sequence, which
CC   can be used in the pharmaceutical composition of the invention. The
CC   invention relates to a pharmaceutical composition containing at least one
CC   activated polyoxyalkylene (APO)-modified cationic peptide. The
CC   modification of peptides with APO increases their activity against tumour
CC   cells, including those with a multidrug resistant phenotype. The
CC   pharmaceutical composition can be used to treat tumours, specifically
CC   lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC   cervix, uterus, skin, prostate, liver and colon.
XX
SQ   Sequence   12 AA;

  Query Match            96.5%;  Score 83;  DB 21;  Length 12;
  Best Local Similarity  91.7%;  Pred. No. 1e-05;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||:|||||||||
Db       1 ilkwpwwpwrrk 12


RESULT  15
```

AAY24594
ID AAY24594 standard; peptide; 12 AA.
XX
AC AAY24594;
XX
DT 18-AUG-1999 (first entry)
XX
DE Indolicidin analogue #46.
XX
KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW additive; shampoo; soap; insecticide; herbicide; preservative;
KW food; technical material.
XX
OS Synthetic.
XX
PN WO9807745-A2.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-US14779.
XX
PR 13-JAN-1997; 97US-0034949.
PR 21-AUG-1996; 96US-0024754.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Erfle D, Fraser JR, Krieger TJ, Taylor R, West MH;
XX
DR WPI; 1998-169090/15.
XX
PT New indolicidin analogues with antimicrobial activity and related
PT nucleic acid - vectors, transformed cells and antibodies, also
PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT toxicity, useful therapeutically, as disinfectants etc.
XX
PS Claim 14; Page 89; 129pp; English.
XX
CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC (II), BBBXZXXZXB (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC (V), LBBnXZnXXZnXRK (VI), LKnXZXXZXRRK (VII) and BBXZXXZXBBB (VIII).
CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC infections caused by bacteria (Gram positive or negative, or anaerobic);
CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC trematodes) or viruses. Typical of very many pathogens that can be
CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC derived from the analogues may be used similarly; the compounds may
CC also be prepared from antibiotics or antiarrythmic agents. The analogues
CC may be used therapeutically or to coat medical devices; also they are
CC useful as surface disinfectants, as additives to shampoo or soaps, as
CC insecticides or herbicides, or as preservatives for foods and technical
CC materials. The analogues are administered by injection, lavage, orally
CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC spectrum of activity than indolicidin and modification as compounds
CC reduces their toxicity.
XX
SQ Sequence 12 AA;

Query Match 95.3%; Score 82; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 LRWPWWPWRRK 12
          ||||||||||
Db      2 lrwpwwpwrrk 12
```

Search completed: January 4, 2002, 08:40:26
Job time: 109 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:40:57 ; Search time 24.75 Seconds
(without alignments)
10.911 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRWPWWPWRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	100.0	12	4	US-08-915-314-42	Sequence 42, Appl
2	86	100.0	20	4	US-08-915-314-47	Sequence 47, Appl
3	86	100.0	21	4	US-08-915-314-46	Sequence 46, Appl
4	86	100.0	21	4	US-08-915-314-48	Sequence 48, Appl
5	83	96.5	12	4	US-08-915-314-40	Sequence 40, Appl
6	82	95.3	12	4	US-08-915-314-76	Sequence 76, Appl
7	81	94.2	12	4	US-08-915-314-77	Sequence 77, Appl
8	81	94.2	12	4	US-08-915-314-87	Sequence 87, Appl
9	80	93.0	12	4	US-08-915-314-78	Sequence 78, Appl
10	80	93.0	12	4	US-08-915-314-85	Sequence 85, Appl
11	80	93.0	12	4	US-08-915-314-86	Sequence 86, Appl
12	78	90.7	12	4	US-08-915-314-80	Sequence 80, Appl
13	78	90.7	12	4	US-08-915-314-83	Sequence 83, Appl
14	78	90.7	13	4	US-08-915-314-38	Sequence 38, Appl
15	77	89.5	12	4	US-08-915-314-69	Sequence 69, Appl
16	75	87.2	11	4	US-08-915-314-41	Sequence 41, Appl
17	75	87.2	12	4	US-08-915-314-52	Sequence 52, Appl
18	75	87.2	13	4	US-08-915-314-25	Sequence 25, Appl
19	75	87.2	13	4	US-08-915-314-30	Sequence 30, Appl
20	75	87.2	13	4	US-08-915-314-51	Sequence 51, Appl
21	75	87.2	13	4	US-08-915-314-62	Sequence 62, Appl
22	75	87.2	13	4	US-08-915-314-63	Sequence 63, Appl
23	75	87.2	13	4	US-08-915-314-64	Sequence 64, Appl
24	75	87.2	13	4	US-08-702-054B-33	Sequence 33, Appl
25	75	87.2	13	4	US-08-702-054B-34	Sequence 34, Appl
26	75	87.2	13	4	US-08-702-054B-35	Sequence 35, Appl
27	75	87.2	13	4	US-09-042-071-36	Sequence 36, Appl
28	75	87.2	14	4	US-08-915-314-57	Sequence 57, Appl
29	75	87.2	16	4	US-08-702-054B-2	Sequence 2, Appli
30	75	87.2	21	4	US-08-915-314-54	Sequence 54, Appl
31	73	84.9	9	4	US-08-915-314-90	Sequence 90, Appl
32	73	84.9	11	4	US-08-915-314-44	Sequence 44, Appl
33	73	84.9	16	4	US-08-702-054B-38	Sequence 38, Appl
34	72	83.7	12	4	US-08-915-314-79	Sequence 79, Appl
35	72	83.7	12	4	US-08-915-314-81	Sequence 81, Appl
36	72	83.7	12	4	US-08-915-314-82	Sequence 82, Appl
37	72	83.7	12	4	US-08-915-314-84	Sequence 84, Appl
38	70	81.4	9	4	US-09-076-227-5	Sequence 5, Appli
39	70	81.4	10	4	US-09-076-227-4	Sequence 4, Appli
40	70	81.4	11	4	US-08-702-054B-9	Sequence 9, Appli
41	70	81.4	11	4	US-09-076-227-3	Sequence 3, Appli
42	70	81.4	12	4	US-08-915-314-39	Sequence 39, Appl
43	70	81.4	12	4	US-08-915-314-74	Sequence 74, Appl
44	70	81.4	12	4	US-08-702-054B-5	Sequence 5, Appli
45	70	81.4	12	4	US-09-076-227-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT   1
US-08-915-314-42
; Sequence 42, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-42


  Query Match             100.0%;  Score 86;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
```

```
          ||||||||||||
Db      1 ILRWPWWPWRRK 12

RESULT   2
US-08-915-314-47
; Sequence 47, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-47


  Query Match             100.0%;  Score 86;  DB 4;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILRWPWWPWRRK 12
          ||||||||||||
Db      1 ILRWPWWPWRRK 12


RESULT   3
US-08-915-314-46
; Sequence 46, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-46


  Query Match             100.0%;  Score 86;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILRWPWWPWRRK 12
          ||||||||||||
Db      1 ILRWPWWPWRRK 12


RESULT   4
US-08-915-314-48
; Sequence 48, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  48:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  21 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-48


  Query Match             100.0%;  Score 86;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||||||||||||
Db       1 ILRWPWWPWRRK 12


RESULT   5
US-08-915-314-40
; Sequence 40, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;     APPLICANT:  Fraser, Janet R.
;     APPLICANT:  West, Michael H.P.
;     APPLICANT:  Krieger, Timothy J.
;     APPLICANT:  Taylor, Robert
;     APPLICANT:  Erfle, Douglas
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;     TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/915,314
;       FILING DATE:  20-AUG-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6180604tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  660081.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  40:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  12 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-40


  Query Match             96.5%;  Score 83;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 3.5e-06;
  Matches   11;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||:|||||||||
Db       1 ILKWPWWPWRRK 12


RESULT   6
US-08-915-314-76
; Sequence 76, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;     APPLICANT:  Fraser, Janet R.
;     APPLICANT:  West, Michael H.P.
;     APPLICANT:  Krieger, Timothy J.
;     APPLICANT:  Taylor, Robert
;     APPLICANT:  Erfle, Douglas
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;     TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/915,314
;       FILING DATE:  20-AUG-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6180604tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  660081.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  76:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  12 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-76


  Query Match             95.3%;  Score 82;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-06;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LRWPWWPWRRK 12
           ||||||||||||
Db       2 LRWPWWPWRRK 12


RESULT   7
US-08-915-314-77
; Sequence 77, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;     APPLICANT:  Fraser, Janet R.
;     APPLICANT:  West, Michael H.P.
;     APPLICANT:  Krieger, Timothy J.
;     APPLICANT:  Taylor, Robert
;     APPLICANT:  Erfle, Douglas
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;     TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
```

```
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  77:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-77


  Query Match             94.2%;  Score 81;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 6.5e-06;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           | ||||||||||
Db       1 IARWPWWPWRRK 12


RESULT   8
US-08-915-314-87
; Sequence 87, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  87:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-87


  Query Match             94.2%;  Score 81;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-06;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRR 11
           |||||||||||
Db       1 ILRWPWWPWRR 11


RESULT   9
US-08-915-314-78
; Sequence 78, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-78


  Query Match             93.0%;  Score 80;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 9e-06;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           || |||||||||
Db       1 ILAWPWWPWRRK 12


RESULT  10
US-08-915-314-85
```

```
; Sequence 85, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  85:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-85


  Query Match             93.0%;  Score 80;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 9e-06;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           |||||||||  ||
Db       1 ILRWPWWPWARK 12


RESULT  11
US-08-915-314-86
; Sequence 86, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  86:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-86


  Query Match             93.0%;  Score 80;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 9e-06;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           |||||||||  |
Db       1 ILRWPWWPWRAK 12


RESULT  12
US-08-915-314-80
; Sequence 80, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  80:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
```

```
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-80


  Query Match             90.7%;  Score 78;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 1.7e-05;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           |||| |||||||
Db       1 ILRWAWWPWRRK 12


RESULT  13
US-08-915-314-83
; Sequence 83, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;     APPLICANT:  Fraser, Janet R.
;     APPLICANT:  West, Michael H.P.
;     APPLICANT:  Krieger, Timothy J.
;     APPLICANT:  Taylor, Robert
;     APPLICANT:  Erfle, Douglas
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;     TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/915,314
;       FILING DATE:  20-AUG-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6180604tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  660081.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  83:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  12 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-83


  Query Match             90.7%;  Score 78;  DB 4;  Length 12;
  Best Local Similarity   91.7%;  Pred. No. 1.7e-05;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           ||||||| ||||
Db       1 ILRWPWWAWRRK 12


RESULT  14
US-08-915-314-38
; Sequence 38, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;     APPLICANT:  Fraser, Janet R.
;     APPLICANT:  West, Michael H.P.
;     APPLICANT:  Krieger, Timothy J.
;     APPLICANT:  Taylor, Robert
;     APPLICANT:  Erfle, Douglas
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;     TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/915,314
;       FILING DATE:  20-AUG-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6180604tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  660081.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  38:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  13 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-38


  Query Match             90.7%;  Score 78;  DB 4;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWPWWPWRRK 12
           ||||||||||
Db       4 RWPWWPWRRK 13


RESULT  15
US-08-915-314-69
; Sequence 69, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;     APPLICANT:  Fraser, Janet R.
;     APPLICANT:  West, Michael H.P.
;     APPLICANT:  Krieger, Timothy J.
;     APPLICANT:  Taylor, Robert
;     APPLICANT:  Erfle, Douglas
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;     TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;     NUMBER OF SEQUENCES:  90
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO: 69:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-69


  Query Match             89.5%;  Score 77;  DB 4;  Length 12;
  Best Local Similarity   83.3%;  Pred. No. 2.3e-05;
  Matches   10;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           | :|||||||||
Db       1 IKKWPWWPWRRK 12


Search completed: January  4, 2002, 08:40:57
Job time: 140 sec
```

```
OM of: US-09-444-281-36 to: N_Geneseq_1101:*   out_format : pfs

Date: Jan 4, 2002 11:02 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084144_16249/app_query.fasta_1.210
 -DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09444281_@CGN1_1_267 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-36
Query length: 12
Database: N_Geneseq_1101:*
Database sequences: 930621
Database length: 428662619
Search time (sec): 286.760000


score_list:
Sequence                    Strd Orig.  ZScore   EScore Len  ! Documentation
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27296 +   86.00  197.06  0.0067    108  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27298 +   86.00  196.62  0.0071    114  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27294 +   86.00  194.32  0.0095    151  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV60908 +   75.00  175.55  0.1058     88  ! C
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27291 +   75.00  173.43  0.1388    114  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV83788 +   70.00  171.67  0.1742     39  ! A
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29389 +   70.00  170.14  0.2118     47  ! F
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29390 -   70.00  170.14  0.2118     47  ! F
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29364 +   70.00  169.16  0.2403     53  ! A
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ40246 +   70.00  167.00  0.3169     69  ! C
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ49764 +   70.00  157.86  1.02       211  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ45123 +   70.00  157.86  1.02       211  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ20646 +   70.00  129.90  36.92     6446  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF82334 +   70.00  129.90  36.92     6446  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28519 +   67.00  151.69  2.26       207  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28510 +   65.00  161.13  0.6731     39  ! R
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28512 -   65.00  158.77  0.9101     52  ! R
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28514 -   65.00  158.77  0.9101     52  ! R
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27290 -   62.00  153.10  1.88       48  ! A
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27287 +   62.00  152.94  1.92       49  ! C
/SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAQ99309 -   61.00  128.43  44.57     758  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ46155 -   61.00  121.95  102.43   1676  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ46160 -   61.00  120.06  130.41   2110  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT21968 +   59.00  130.42  34.55      355  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX84205 +   59.00  129.82  37.31      382  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC79434 +   59.00  129.82  37.31      382  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV61297 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV58559 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA06322 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAS10081 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH93438 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH84752 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH02503 +   59.00  129.80  37.41      383  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV61328 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV58653 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA06416 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAS10175 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH93532 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH84846 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH02597 +   59.00  129.78  37.51      384  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX84191 +   59.00  128.67  43.27      440
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC79420 +   59.00  128.67  43.27      440
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV61188 +   59.00  127.64  49.37      499
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV58550 +   59.00  127.64  49.37      499
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA06313 +   59.00  127.64  49.37      499

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27296

seq_documentation_block:
ID   AAA27296 standard; DNA; 108 BP.
XX
AC   AAA27296;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI 2X11B7 poly cassette.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 5; Page 39; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11B7 fragment. This fragment was used in the expression
CC   cassette. MBI-11B7 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 108 BP; 18 A; 33 C; 31 G; 26 T; 0 other;


alignment_scores:
           Quality:   86.00              Length:     12
             Ratio:  7.167                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-36 x AAA27296   ..

 Align seg 1/1  to: AAA27296  from: 1  to: 108

       1 IleLeuArgTrpProTrpTrpProTrpArgArgLys 12
         ||||||||||||||||||||||||||||||||||||
      41 ATTCTGCGTTGGCCGTGGTGGCCGTGGCGTCGCAAA 76

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27298
```

```
seq_documentation_block:
ID   AAA27298 standard; DNA; 114 BP.
XX
AC   AAA27298;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI 11B7 first cassette.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 5; Page 40; [illegible]; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11B7 fragment. This fragment was used in the expression
CC   cassette. MBI-11B7 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 114 BP; 20 A; 34 C; 32 G; 28 T; 0 other;


alignment_scores:
            Quality:  86.00              Length:     12
              Ratio:  7.167                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-36 x AAA27298   ..

 Align seg 1/1  to: AAA27298  from: 1  to: 114

       1 IleLeuArgTrpProTrpTrpProTrpArgArgLys 12
         ||||||||||||||||||||||||||||||||||||
      47 ATTCTGCGTTGGCCGTGGTGGCCGTGGCGTCGCAAA 82

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27294

seq_documentation_block:
ID   AAA27294 standard; DNA; 151 BP.
XX
AC   AAA27294;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI 2X11B7 last cassette.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 5; Page 38; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11B7 fragment. This fragment was used in the expression
CC   cassette. MBI-11B7 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 151 BP; 22 A; 44 C; 49 G; 36 T; 0 other;


alignment_scores:
            Quality:  86.00              Length:     12
              Ratio:  7.167                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-36 x AAA27294   ..

 Align seg 1/1  to: AAA27294  from: 1  to: 151

       1 IleLeuArgTrpProTrpTrpProTrpArgArgLys 12
         ||||||||||||||||||||||||||||||||||||
      41 ATTCTGCGTTGGCCGTGGTGGCCGTGGCGTCGCAAA 76

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV60908

seq_documentation_block:
ID   AAV60908 standard; DNA; 88 BP.
XX
AC   AAV60908;
XX
DT   11-JAN-1999  (first entry)
XX
DE   DNA fragment encoding MBI11.
XX
KW   MBI28; cationic peptide; plasmid pKL1; small cryptic plasmid;
KW   replication; RepA; vector; RAMP; human; MBI11; ss.
XX
OS   Synthetic.
```

```
OS   Homo sapiens.
XX
PN   WO9841636-A2.
XX
PD   24-SEP-1998.
XX
PF   16-MAR-1998;   98WO-CA00214.
XX
PR   14-MAR-1997;   97US-0040722.
XX
PA   (BURI/) BURIAN J.
PA   (KAYW/) KAY W W.
XX
PI   Burian J,  Kay WW;
XX
DR   WPI; 1998-531571/45.
XX
PT   Increasing plasmid copy number in a cell with the repA gene product
PT   - and an small cryptic plasmid ori sequence, useful for high level
PT   expression of e.g. cytokines, antigens or therapeutic proteins
XX
PS   Example 16; Page 57; 82pp; English.
XX
CC   This oligonucleotide was used as a template in a PCR reaction (see
CC   also AAV60909-10) to generate a DNA fragment encoding the cationic
CC   peptide MBI11 (see AAW71690).  The PCR product was cloned into
CC   the universal vector pR2h-B1, which contains the R21 replication
CC   leader of RepA (see AAW71686) and 2 tandem copies of the prepro
CC   region (Hpro) of human defensin.  The vector provides expression
CC   of R21-hpro-MBI11 fusion in host cells.  The invention provides
CC   controlled replication plasmid vectors (RAMP vectors) comprising a
CC   replication origin of a small cryptic plasmid such as pKL1 (see
CC   AAV58292) and a gene encoding RepA (see AAW71686).  The vectors can
CC   reach very high levels of plasmid replication, but are not lethal
CC   to the host cell, and can be used to direct the high level
CC   expression of e.g. cytokines, antigens and therapeutic proteins.
XX
SQ   Sequence 88 BP; 20 A; 18 C; 25 G; 25 T; 0 other;


alignment_scores:
            Quality:   75.00              Length:     10
              Ratio:  7.500                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
 US-09-444-281-36 x AAV60908    ..

 Align seg 1/1  to: AAV60908  from: 1  to: 88

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         :::||||||||||||||||||||||||||||
      34 AAATGGCCGTGGTGGCCGTGGCGTCGTAAA 63

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27291

seq_documentation_block:
ID   AAA27291 standard; DNA; 114 BP.
XX
AC   AAA27291;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI-11 fragment.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 4; Page 37; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11 fragment. MBI-11 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 114 BP; 25 A; 26 C; 30 G; 33 T; 0 other;


alignment_scores:
            Quality:   75.00              Length:     10
              Ratio:  7.500                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
 US-09-444-281-36 x AAA27291    ..

 Align seg 1/1  to: AAA27291  from: 1  to: 114

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         :::||||||||||||||||||||||||||||
      50 AAATGGCCGTGGTGGCCGTGGCGTCGTAAA 79

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV83788

seq_documentation_block:
ID   AAV83788 standard; DNA; 39 BP.
XX
AC   AAV83788;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Antimicrobial peptide Indolicidin encoding DNA.
XX
KW   Antimicrobial; fusion; acidic peptide; recombinant; microorganism;
KW   guamerin; basic peptide; Indolicidin; ss.
XX
OS   Synthetic.
OS   Bos sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..39
FT                   /*tag=  a
FT                   /note= "the start and stop codons are not indicated"
XX
PN   WO9854336-A1.
XX
PD   03-DEC-1998.
XX
PF   28-MAY-1998;   98WO-KR00132.
```

XX
PR 09-APR-1998; 98KR-0013372.
PR 28-MAY-1997; 97KR-0021312.
XX
PA (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
PA (SAMY-) SAMYANG GENEX CORP.
XX
PI Hong S, Kang MH, Kim JH, Kim S, Lee H, Lee JH;
XX
DR WPI; 1999-059844/05.
DR P-PSDB; AAW87609.
XX
PT New method for mass production of antimicrobial peptides - by
PT constructing fusion genes comprising acidic and antimicrobial
PT peptide genes and transforming host with vector containing these
XX
PS Example 6; Page 18; 52pp; English.
XX
CC The invention relates to mass production of antimicrobial peptides. The
CC method comprises constructing a fusion gene containing a first gene
CC encoding a negatively charged acidic peptide having at least two cysteine
CC residues, and a second gene encoding a positively charged basic
CC antimicrobial peptide. A host microorganism is transformed with a vector
CC containing the fusion gene and then cultured. The expressed antimicrobial
CC peptide is then recovered. The method is used to mass produce
CC antimicrobial peptides in recombinant microorganisms. The inhibitory
CC effect of the expressed antimicrobial peptide upon the growth of the host
CC microorganism is considerably reduced by fusing it to the acidic peptide.
CC Therefore, the use of the fusion gene provides an economic, recombinant
CC alternative of mass producing antimicrobial peptides, which overcomes the
CC disadvantages of low-productivity and poor economy, previously
CC encountered by recombinant and chemical methods. The present sequence
CC represents the DNA encoding an antimicrobial peptide Indolicidin. This
CC can be used along with the acidic peptide Guamerin gene in the
CC construction of the fusion gene.
XX
SQ Sequence 39 BP; 4 A; 10 C; 16 G; 9 T; 0 other;

alignment_scores:
Quality: 70.00 Length: 9
Ratio: 7.778 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-444-281-36 x AAV83788 ..

Align seg 1/1 to: AAV83788 from: 1 to: 39

3 ArgTrpProTrpTrpProTrpArgArg 11
::::||||||||||||||||||||||||
13 AAATGGCCGTGGTGGCCGTGGCGTCGT 39

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29389

seq_documentation_block:
ID AAZ29389 standard; DNA; 47 BP.
XX
AC AAZ29389;
XX
DT 29-FEB-2000 (first entry)
XX
DE PCR primer-15 for synthesis of antimicrobial peptide Indolicidin gene.
XX
KW PCR primer; anti-microbial peptide; Indolicidin gene; DNA construct;
KW glutamine pyrophosphoribosyl pyrophosphate amidotransferase gene;
KW purF gene; fusion peptide; mass production; pharmaceutical industry;
KW food industry; ss.
XX
OS Synthetic.
XX
PN WO9964611-A1.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-KR00282.
XX
PR 09-JUN-1998; 98KR-0022117.
PR 14-MAY-1999; 99KR-0017920.
XX
PA (SAMY-) SAMYANG GENEX CORP.
XX
PI Kim JH, Kang MH, Lee J, Park SH, Lee JW, Hong SS, Lee H;
XX
DR WPI; 2000-097542/08.
XX
PT New DNA constructs useful for mass production of antimicrobial peptides
PT in microorganism hosts -
XX
PS Example 1; Page 13; 67pp; English.
XX
CC The present sequence is a chemically synthesised PCR primer which was
CC used to synthesise a gene encoding antimicrobial peptide Indolicidin.
CC The antimicrobial peptide gene is used in a DNA construct that comprises
CC entire, partial or a derivative of purF gene (glutamine
CC pyrophosphoribosyl pyrophosphate amidotransferase gene). The DNA
CC construct allows mass production of the antimicrobial peptide in
CC microbial hosts without killing the host cells. The antimicrobial
CC peptides are useful commercially in the pharmaceutical and
CC food industries.
XX
SQ Sequence 47 BP; 6 A; 11 C; 19 G; 11 T; 0 other;

alignment_scores:
Quality: 70.00 Length: 9
Ratio: 7.778 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-444-281-36 x AAZ29389 ..

Align seg 1/1 to: AAZ29389 from: 1 to: 47

3 ArgTrpProTrpTrpProTrpArgArg 11
::::||||||||||||||||||||||||
17 AAATGGCCGTGGTGGCCGTGGCGTCGT 43

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29390

seq_documentation_block:
ID AAZ29390 standard; DNA; 47 BP.
XX
AC AAZ29390;
XX
DT 29-FEB-2000 (first entry)
XX
DE PCR primer-16 for synthesis of antimicrobial peptide Indolicidin gene.
XX
KW PCR primer; anti-microbial peptide; Indolicidin gene; DNA construct;
KW glutamine pyrophosphoribosyl pyrophosphate amidotransferase gene;
KW purF gene; fusion peptide; mass production; pharmaceutical industry;
KW food industry; ss.
XX
OS Synthetic.
XX
PN WO9964611-A1.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-KR00282.
XX
PR 09-JUN-1998; 98KR-0022117.
PR 14-MAY-1999; 99KR-0017920.
XX
PA (SAMY-) SAMYANG GENEX CORP.

```
XX
PI   Kim JH,  Kang MH,  Lee J,  Park SH,  Lee JW,  Hong SS,  Lee H;
XX
DR   WPI; 2000-097542/08.
XX
PT   New DNA constructs useful for mass production of antimicrobial peptides
PT   in microorganism hosts  -
XX
PS   Example 1; Page 13; 67pp; English.
XX
CC   The present sequence is a chemically synthesised PCR primer which was
CC   used to synthesise a gene encoding antimicrobial peptide Indolicidin.
CC   The antimicrobial peptide gene is used in a DNA construct that comprises
CC   entire, partial or a derivative of purF gene (glutamine
CC   pyrophosphoribosyl pyrophosphate amidotransferase gene). The DNA
CC   construct allows mass production of the antimicrobial peptide in
CC   microbial hosts without killing the host cells. The antimicrobial
CC   peptides are useful commercially in the pharmaceutical and
CC   food industries.
XX
SQ   Sequence 47 BP; 12 A; 18 C; 10 G; 7 T; 0 other;


alignment_scores:
            Quality:   70.00              Length:      9
              Ratio:  7.778                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AAZ29390/rev   ..

 Align seg 1/1  to reverse of: AAZ29390  from: 1  to: 47

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
      35 AAATGGCCGTGGTGGCCGTGGCGTCGT 9

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29364

seq_documentation_block:
ID   AAZ29364 standard; DNA; 53 BP.
XX
AC   AAZ29364;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Antimicrobial peptide, Indolicidin encoding DNA.
XX
KW   purF gene; glutamine pyrophosphoribosyl pyrophosphate amidotransferase;
KW   purF derivative; fusion partner; antimicrobial peptide; Indolicidin;
KW   mass production; cleavage site; hydroxylamine; CNBr; DNA construct; cow;
KW   neutralise; toxicity; pharmaceutical industry; food industry; ds.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             5..46
FT                   /*tag=  a
FT                   /product= "Indolicidin peptide"
FT                   /note= "Antimicrobial peptide used in DNA construct"
XX
PN   WO9964611-A1.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999;   99WO-KR00282.
XX
PR   09-JUN-1998;   98KR-0022117.
PR   14-MAY-1999;   99KR-0017920.
XX
PA   (SAMY-) SAMYANG GENEX CORP.
XX
PI   Kim JH,  Kang MH,  Lee J,  Park SH,  Lee JW,  Hong SS,  Lee H;
XX
DR   WPI; 2000-097542/08.
DR   P-PSDB; AAY44324.
XX
PT   New DNA constructs useful for mass production of antimicrobial peptides
PT   in microorganism hosts  -
XX
PS   Claim 1; Fig 1; 67pp; English.
XX
CC   The present DNA sequence encodes an antimicrobial peptide, Indolicidin
CC   derived from cow, Bos taurus. It is used along with a derivative
CC   of purF gene sequence that functions as a fusion partner.
CC   A DNA construct that comprises, this antimicrobial peptide encoding
CC   sequence and the entire, partial or derivative of purF gene, is used for
CC   mass production of the antimicrobial peptide in microorganisms without
CC   killing the host cells. Use of the purF gene derivative sequence,
CC   neutralises the toxicity of the antimicrobial peptides against the host
CC   microorganism. The antimicrobial peptides are useful commercially in the
CC   pharmaceutical and food industries.
XX
SQ   Sequence 53 BP; 8 A; 12 C; 20 G; 13 T; 0 other;


alignment_scores:
            Quality:   70.00              Length:      9
              Ratio:  7.778                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AAZ29364   ..

 Align seg 1/1  to: AAZ29364  from: 1  to: 53

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
      17 AAATGGCCGTGGTGGCCGTGGCGTCGT 43

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ40246

seq_documentation_block:
ID   AAZ40246 standard; DNA; 69 BP.
XX
AC   AAZ40246;
XX
DT   23-FEB-2000  (first entry)
XX
DE   Oligonucleotide for cloning indolicidin peptide coding sequence.
XX
KW   Indolicidin; bactenecin; sulphate-reducing bacteria; growth inhibitor;
KW   corrosion; degradation; metal; concrete; cement; dental implant; biofilm;
KW   ss.
XX
OS   Synthetic.
OS   Bacillus sp.
XX
PN   WO9956553-A1.
XX
PD   11-NOV-1999.
XX
PF   03-MAY-1999;   99WO-US09675.
XX
PR   06-MAY-1998;   98US-0074037.
PR   31-MAR-1999;   99US-0282277.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Wood TK,  Jayaraman A,  Earthman JC;
XX
DR   WPI; 2000-052882/04.
XX
PT   Inhibiting growth of sulphate-reducing bacteria using other bacteria,
PT   particularly for protection of metals and concrete  -
XX
```

```
PS   Example 4; Fig 1; 84pp; English.
XX
CC   This sequence represents an oligonucleotide for cloning the non-amidated
CC   indolicidin peptide coding sequence. The invention relates to a method
CC   for inhibiting growth of sulphate-reducing bacteria (A) on a material (B)
CC   sensitive to corrosion or degradation, by applying to (B) a bacterium (C)
CC   that secretes a compound (I) able to inhibit growth of (A). The method is
CC   used to protect metal, concrete or cement against corrosion and
CC   degradation, but (B) can also be used to protect dental implants. (B) is
CC   present in an open or closed system (e.g. water cooling tower, liquid
CC   storage container, fuel tank, sewer or drainage system etc.) or part of a
CC   bridge or other structure. The method is more effective and less
CC   expensive than known methods for inhibiting (A), and reduces the amount
CC   of toxic chemicals released. Conventional biofilms of aerobic organisms
CC   tend to encourage growth of (A), and addition of (C) to the biofilm
CC   prevents this. A single application of (C) lasts for a long time, and (I)
CC   are produced exactly where they are required and inhibit (A) without
CC   significant impact on other organisms (this effect includes reducing
CC   resistance of (A) to conventional biocides, which may then be used in
CC   reduced amounts). If local damage to the biofilm occurs, the underlying
CC   material is still protected by diffusion of (I) from neighbouring areas.
XX
SQ   Sequence 69 BP; 14 A; 18 C; 20 G; 17 T; 0 other;


alignment_scores:
           Quality:   70.00              Length:      9
             Ratio:  7.778                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AAZ40246    ..

 Align seg 1/1  to: AAZ40246  from: 1  to: 69

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
      28 AAATGGCCTTGGTGGCCTTGGCGCCGC 54

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ49764

seq_documentation_block:
ID   AAZ49764 standard; DNA; 211 BP.
XX
AC   AAZ49764;
XX
DT   18-APR-2000  (first entry)
XX
DE   Poly-(Indol (1-13)-Met-Ala-Arg-Ile-Ala-Met)3 DNA.
XX
KW   Crosslinked indolicidin analog; X-indolicidin; poly-Indol 1-13;
KW   stability; bovine neutrophil; antimicrobial; antibacterial; fungicide;
KW   protozoacide; virucide; anti-HIV; human immunodeficiency virus-1;
KW   HIV-1; gram positive bacteria; gram negative; Staphylococcus aureus;
KW   Escherichia coli; Salmonella typhimurium; yeast; fungi; protozoa;
KW   Candida albicans; Cryptococcus neoformans; Giardia; Acanthamoeba;
KW   hexapeptide spacer; ds.
XX
OS   Synthetic.
OS   Bos sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             8..199
FT                   /*tag=  a
FT                   /product= "Poly-(Indol(1-13)-Met-Ala-Arg-Ile-Ala-Met)3"
FT                   /note= "encodes three copies of Indol 1-13, each
FT                   separated by Met-Ala-Arg-Ile-Ala-Met spacer sequence"
FT   primer_bind     1..21
FT                   /*tag=  b
FT   primer_bind     complement (191..211)
FT                   /*tag=  c
FT   misc_feature    68..71
FT                   /*tag=  d
FT                   /note= "corresponds to overlap in oligonucleotides
FT                   used for ligation"
FT   misc_feature    148..151
FT                   /*tag=  e
FT                   /note= "corresponds to overlap in oligonucleotides
FT                   used for ligation"
XX
PN   WO9965510-A1.
XX
PD   23-DEC-1999.
XX
PF   20-MAY-1999;   99WO-US11165.
XX
PR   18-JUN-1998;   98US-0099631.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Selsted ME,  Osapay K;
XX
DR   WPI; 2000-147133/13.
DR   P-PSDB; AAY44668.
XX
PT   Crosslinked indolicidin analogs with antimicrobial activity against
PT   bacteria, yeast, fungi, protozoa and viruses  -
XX
PS   Example 1C; Fig 1; 53pp; English.
XX
CC   The patent discloses crosslinked analogs of indolicidin (Indol 1-13)
CC   which is a naturally occurring peptide isolated from bovine neutrophils
CC   and has antimicrobial activity. The crosslinked indolicidin
CC   (X-indolicidin) analogs are stable and have antimicrobial activity
CC   against gram positive and negative bacteria (e.g. Staphylococcus aureus,
CC   Escherichia coli and Salmonella typhimurium), yeasts and fungi (e.g.
CC   Candida albicans, Cryptococcus neoformans), protozoa (e.g. Giardia
CC   species and Acanthamoeba species), and viruses (e.g. HIV-1).
CC   They can be used for reducing or inhibiting the growth or survival of
CC   microorganisms in an environment e.g. a food or food product, a
CC   solution, an inanimate object comprising a surface, or a mammal.
CC   The present sequence is a DNA encoding a protein comprising three
CC   copies of Indol 1-13 each separated by a hexapeptide spacer sequence.
CC   The sequence was used to produce a recombinant construct for the
CC   expression of Indol-homoserine (Hse) analog. The ability of
CC   Indol-Hse analog to maintain antimicrobial activity provides a means to
CC   produce X-indolicidin analog precursors in sufficient quantities.
XX
SQ   Sequence 211 BP; 36 A; 50 C; 74 G; 51 T; 0 other;


alignment_scores:
           Quality:   70.00              Length:      9
             Ratio:  7.778                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AAZ49764    ..

 Align seg 1/1  to: AAZ49764  from: 1  to: 211

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
      38 AAATGGCCGTGGTGGCCGTGGCGTCGT 64

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ45123

seq_documentation_block:
ID   AAZ45123 standard; DNA; 211 BP.
XX
AC   AAZ45123;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Indolicidin fusion peptide nucleotide sequence.
XX
```

KW Indolicidin analogue; antimicrobial activity; helminth; bacteria; virus;
KW treatment; inhibit growth; micro-organism; contact lens solution;
KW transgenic plant; surgical instrument; yeast; fungi; protozoa; ss.
XX
OS Synthetic.
XX
PN WO9958141-A1.
XX
PD 18-NOV-1999.
XX
PF 05-MAY-1999; 99WO-US09942.
XX
PR 12-MAY-1998; 98US-0076227.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Selsted ME;
XX
DR WPI; 2000-053028/04.
DR P-PSDB; AAY57142.
XX
PT New indolicidin analogues, active against bacteria, yeast, fungi,
PT protozoa and virus, used for, e.g. treating infections -
XX
PS Disclosure; Fig 6; 62pp; English.
XX
CC This is the nucleotide sequence of an example of a fusion protein which
CC consists of an indolicidin analogue linked to another peptide.
CC Peptides AAY57109-Y57138 and AAY57143-Y57144 are new indolicidin
CC analogues, which have a homoserine residue and/or a truncated amino
CC terminal region. The analogues have the following amino acid sequence:
CC Xaa1-Xaa2-Xaa3-Xaa4-Xaa5-Xaa6-Pro-Xaa6-Xaa6-Pro-Xaa6-Xaa7-Xaa7-Xaa8
CC Where:
CC Xaa1 = Ile, Leu, Val, Ala, Gly or absent;
CC Xaa2 = Ile, Leu, Val, Ala, Gly or absent;
CC Xaa3 = Pro or absent;
CC Xaa4 = Trp, Phe or absent;
CC Xaa5 = Arg, Lys or absent;
CC Xaa6 = Trp or Phe;
CC Xaa7 = Arg, Lys or absent;
CC Xaa8 = homoserine (Hse), Met, Met-Xaa9-Met or absent, and
CC Xaa9 = at least one amino acid;
CC provided that if Xaa1 is present, Xaa8 = Hse, Met or Met-Xaa9-Met;
CC and further provided that: if Xaa2 is absent, Xaa1 is absent; if Xaa3 is
CC absent, Xaa1 and Xaa2 are absent; if Xaa4 is absent, Xaa1, Xaa2 and Xaa3
CC are absent; and if Xaa5 is absent, Xaa1, Xaa2, Xaa3 and Xaa4 are absent.
CC The indolicidin analogues can be used to create a fusion polypeptide
CC consisting of the analogue linked to a peptide. The indolicidin
CC analogues have antimicrobial activity against gram positive bacteria,
CC gram negative bacteria, yeast, fungus, protozoa and viruses (e.g. HIV-1).
CC They are also active against helminths. The analogues can be used for
CC reducing or inhibiting growth or survival of a microorganism. They can be
CC used for treating infections. They can also be included in a liquid such
CC as water or an aqueous solution, e.g. contact lens solution. The
CC analogues have potential uses in food products, and in objects such as
CC the surface of an instrument used to prepare food or to perform surgery.
CC Transgenic plants or animals useful in the food industry can be produced
CC by introducing a nucleic acid molecule encoding an indolicidin analogue
CC into the germline cells of such organisms.
XX
SQ Sequence 211 BP; 36 A; 50 C; 74 G; 51 T; 0 other;

alignment_scores:
Quality: 70.00 Length: 9
Ratio: 7.778 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-444-281-36 x AAZ45123 ..

Align seg 1/1 to: AAZ45123 from: 1 to: 211

3 ArgTrpProTrpTrpProTrpArgArg 11
::::||||||||||||||||||||||||
38 AAATGGCCGTGGTGGCCGTGGCGTCGT 64

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ20646

seq_documentation_block:
ID AAZ20646 standard; RNA; 6446 BP.
XX
AC AAZ20646;
XX
DT 26-NOV-1999 (first entry)
XX
DE TMV-based virus TMV861 coat protein read-through RNA sequence.
XX
KW TMV-based virus; tobacco mosaic virus; protein isolation; green juice;
KW virus isolation; fusion protein identification; ss.
XX
OS Tobacco mosaic virus.
XX
PN WO9946288-A2.
XX
PD 16-SEP-1999.
XX
PF 09-MAR-1999; 99WO-US05056.
XX
PR 10-MAR-1998; 98US-0037751.
XX
PA (BIOS-) BIOSOURCE TECHNOLOGIES INC.
XX
PI Garger SJ, Holtz RB, McCulloch MJ, Turpen TH;
XX
DR WPI; 1999-561660/47.
XX
PT Obtaining protein, viruses and fusion proteins from plants, using
PT non-denaturing conditions -
XX
PS Disclosure; Page 55-58; 58pp; English.
XX
CC This sequence represents a tobacco mosaic virus (TMV) based virus
CC sequence identified using the method of the invention. The method is for
CC obtaining a soluble protein or peptide of interest from a plant,
CC comprises homogenising the plant to produce green juice, adjusting the
CC pH to less than or equal to 5.2, and heating the juice to a minimum of
CC 45 degrees C. The juice is then centrifuged to produce a supernatant, and
CC the protein or peptide is purified from the supernatant. The method can
CC also be used for obtaining viruses and fusion proteins. The method is
CC especially useful for obtaining IL-1 to IL-10, EPO, G-CSF, GM-CSF,
CC hP-CSF, M-CSF, Factor VIII, Factor IX, tPA, receptors, receptor
CC antagonists, antibodies, single-chain antibodies, enzymes,
CC neuropolypeptides, insulin, antigens, vaccines, peptide hormones,
CC calcitonin, and human growth hormone, or an antimicrobial peptide or
CC protein from protegrins, magainins, cecropins, melittins, indolicidins,
CC defensins, beta-defensins, cryptdins, clavainins, plant defensins,
CC nicin and bactenecins, all produced by recombinant means. The new method
CC is more efficient than the prior art for isolating viruses, protein, and
CC peptides. The method is large-scale, and non-denaturing and
CC solvent-limited. Prior art methods do not isolate recombinant proteins,
CC and do not allow fraction 2 proteins to be ultrafiltrated.
XX
SQ Sequence 6446 BP; 1873 A; 1234 C; 1563 G; 1776 U; 0 other;

alignment_scores:
Quality: 70.00 Length: 9
Ratio: 7.778 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-444-281-36 x AAZ20646 ..

Align seg 1/1 to: AAZ20646 from: 1 to: 6446

```
      3 ArgTrpProTrpTrpProTrpArgArg 11
        :::||||||||||||||||||||||||
   6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF82334

seq_documentation_block:
ID AAF82334 standard; RNA; 6446 BP.
XX
AC AAF82334;
XX
DT 22-JUN-2001 (first entry)
XX
DE Tobacco mosaic virus-based coat protein read-through virus TMV861.
XX
KW TMV; tobacco mosaic virus; TMV861; virus isolation;
KW non-native protein purification; ribulose 1,5-diphosphate carboxylase;
KW RuBisCo; coat protein read-through; ss.
XX
OS Tobacco mosaic virus.
XX
PN WO200119969-A1.
XX
PD 22-MAR-2001.
XX
PF 19-MAY-2000; 2000WO-US13680.
XX
PR 16-SEP-1999; 99US-0397090.
XX
PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX
PI Garger SJ, Holtz BR, McCulloch MJ, Turpen TH;
XX
DR WPI; 2001-328016/34.
XX
PT Minimizing presence of ribulose 1,5-diphosphate carboxylase to obtain
PT plant product for isolating bioactive species involves cutting plant
PT material from plant in cutting period when quantity of RuBisCo is at
PT minimum -
XX
PS Disclosure; Page 72-74; 81pp; English.
XX
CC The present sequence is a tobacco mosaic virus (TMV)-based virus
CC which was used to infect field-grown tobacco. The virus was then
CC isolated from the tobacco plants by a novel process for isolating and
CC purifying viruses, soluble proteins and peptides from plant sources. In
CC order to isolate the bioactive species from the undesirable
CC photosynthetic protein ribulose 1,5-diphosphate carboxylase (RuBisCo),
CC the plant material is cut in a period of the light/dark cycle when the
CC quantity of RuBisCo in the plant is at a minimum. The method is useful
CC for obtaining a virus of interest. It is also useful for obtaining
CC soluble recombinant or non-native proteins, such as active mammalian
CC proteins, enzymes, vaccines, antibodies and peptides, from transgenic
CC plants.
XX
SQ Sequence 6446 BP; 1873 A; 1234 C; 1563 G; 1776 U; 0 other;

alignment_scores:
Quality: 70.00 Length: 9
Ratio: 7.778 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-444-281-36 x AAF82334 ..

Align seg 1/1 to: AAF82334 from: 1 to: 6446

```
      3 ArgTrpProTrpTrpProTrpArgArg 11
        :::||||||||||||||||||||||||||
   6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28519

seq_documentation_block:
ID AAA28519 standard; DNA; 207 BP.
XX
AC AAA28519;
XX
DT 29-AUG-2000 (first entry)
XX
DE PCRIL DNA coding sequence.
XX
KW Magainin; antimicrobial; transgenic plant; protease degradation; Rev4;
KW indolicidin; protein production; reverse peptide; ss.
XX
OS Synthetic.
XX
PN WO200026344-A1.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-US25561.
XX
PR 30-OCT-1998; 98US-0106373.
PR 02-NOV-1998; 98US-0106537.
XX
PA (INTE-) INTERLINK BIOTECHNOLOGIES LLC.
PA (KENT) UNIV KENTUCKY RES FOUND.
XX
PI Everett NP, Li Q, Lawrence C, Davies MH;
XX
DR WPI; 2000-365597/31.
DR P-PSDB; AAY92840.
XX
PT Polypeptides for reducing proteolytic degradation of proteins
PT administered to, or produced by a plant comprise indolicin or its
PT functional equivalents
XX
PS Example 17; Page 35; 50pp; English.
XX
CC Indolicidin is a potent antimicrobial tridecapeptide, originally
CC purified from cytoplasmic granules of bovine neutrophils. Reverse
CC peptide, Rev4 of indolicidin (see AAY92794) was found to have increased
CC stability against plant protease degradation. Expression of antimicrobial
CC peptides in transgenic plants suffers a major limitation in that the
CC foreign peptides are susceptible to rapid degradation by proteases. The
CC invention concerns reducing the extent of protease degradation of a
CC protein applied to, or produced by a plant by administering indolicidin,
CC Rev4 or a functional equivalent to the plant. Transgenic plants
CC expressing indolicidin and Rev4 are useful for production of the
CC antimicrobial peptides. Compositions containing indolicidin and Rev4 are
CC also useful for production of agronomically important proteins in
CC plants.
XX
SQ Sequence 207 BP; 49 A; 50 C; 36 G; 72 T; 0 other;

alignment_scores:
Quality: 67.00 Length: 10
Ratio: 7.444 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:
US-09-444-281-36 x AAA28519 ..

Align seg 1/1 to: AAA28519 from: 1 to: 207

```
      1 IleLeuArgTrpProTrpTrpProTrpArg 10
        |||   ||||||||||||||||||||:::
    163 ATTAGGAGATGGCCTTGGTGGCCTTGGAAA 192
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:40:26 ; Search time 53.46 Seconds
(without alignments)
18.013 Million cell updates/sec

Title: US-09-444-281-35
Perfect score: 91
Sequence: 1 ILKKWPWWPWRRK 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	91	100.0	13	18	AAW12873	Antimicrobial cati
2	91	100.0	13	19	AAY24609	Indolicidin analog
3	91	100.0	13	19	AAW66378	Cationic peptide o
4	91	100.0	13	19	AAW71690	Cationic peptide M
5	91	100.0	13	21	AAY94495	MBI-11 peptide der
6	91	100.0	13	21	AAY92795	Indolicidin analog
7	91	100.0	13	21	AAY91773	Amino acid sequenc
8	91	100.0	13	21	AAY91774	Amino acid sequenc
9	91	100.0	13	21	AAY91818	Amino acid sequenc
10	91	100.0	13	21	AAY91819	Amino acid sequenc
11	91	100.0	13	21	AAY91820	Amino acid sequenc
12	91	100.0	14	19	AAY24583	Indolicidin analog
13	91	100.0	14	21	AAY91811	Amino acid sequenc
14	91	100.0	21	19	AAY24582	Indolicidin analog
15	91	100.0	21	21	AAY91806	Amino acid sequenc
16	87	95.6	12	19	AAY24580	Indolicidin analog
17	87	95.6	12	21	AAY91804	Amino acid sequenc
18	86	94.5	12	18	AAW12877	Antimicrobial cati
19	86	94.5	12	19	AAY24615	Indolicidin analog
20	86	94.5	12	21	AAY91833	Amino acid sequenc
21	86	94.5	13	19	AAY24572	Indolicidin analog
22	86	94.5	13	21	AAY91812	Amino acid sequenc
23	86	94.5	14	19	AAY24573	Indolicidin analog
24	86	94.5	14	21	AAY91813	Amino acid sequenc
25	86	94.5	15	18	AAW13802	Antimicrobial cati
26	86	94.5	20	19	AAY24570	Indolicidin analog
27	86	94.5	20	21	AAY91807	Amino acid sequenc
28	86	94.5	21	19	AAY24571	Indolicidin analog
29	86	94.5	21	21	AAY91808	Amino acid sequenc
30	85	93.4	12	19	AAY24586	Indolicidin analog
31	85	93.4	12	21	AAY91828	Amino acid sequenc
32	85	93.4	13	18	AAW27179	Antimicrobial cati
33	85	93.4	13	18	AAW12889	Antimicrobial cati
34	85	93.4	13	18	AAW12894	Antimicrobial cati
35	85	93.4	13	19	AAY24610	Indolicidin analog
36	85	93.4	13	19	AAY24565	Indolicidin analog
37	85	93.4	13	21	AAY91786	Amino acid sequenc
38	85	93.4	13	21	AAY91795	Amino acid sequenc
39	83	91.2	12	19	AAY24568	Indolicidin analog
40	83	91.2	12	21	AAY91789	Amino acid sequenc
41	83	91.2	13	18	AAW12892	Antimicrobial cati
42	83	91.2	13	18	AAW12893	Antimicrobial cati
43	83	91.2	13	18	AAW12896	Antimicrobial cati
44	83	91.2	13	18	AAW12897	Antimicrobial cati
45	83	91.2	13	19	AAY24612	Indolicidin analog

ALIGNMENTS

RESULT 1
AAW12873
ID AAW12873 standard; peptide; 13 AA.
XX
AC AAW12873;
XX
DT 10-DEC-1997 (first entry)
XX
DE Antimicrobial cationic peptide CP-11.
XX
KW Bacterial; viral; antitumour; food; preservative; inhibitor; growth;
KW bacterium; yeast; endotoxaemia; sepsis; antibiotic; fungal;
KW antiviral; Candida albicans; sterilant; Salmonella; Yersina;
KW Shigella.
XX
OS Synthetic.
XX
PN WO9708199-A2.
XX
PD 06-MAR-1997.
XX
PF 23-AUG-1996; 96WO-IB00996.
XX
PR 23-AUG-1995; 95US-0002687.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Falla TJ, Gough M, Hancock REW;
XX
DR WPI; 1997-179179/16.
XX
PT Cationic peptide(s) having anti-microbial activity - used for the
PT inhibition of bacterial and viral growth, as an antitumour agent,
PT and as a food preservative

XX
PS Claim 2; Page 65; 89pp; English.
XX
CC The present sequence represents a specifically claimed novel isolated
CC cationic peptide which has antimicrobial activity. The amino acid
CC sequence of antimicrobial cationic peptides (including the present
CC sequence) is selected from: X1X1ProX2X3X2Pro(X2X2Pro)nX2X3(X5)o;
CC X1X1ProX2X3X4(X5)rProX2X3X3; X1X1X3(ProTrp)uX3X2X5X2X2X5X2(X5)o;
CC X1X1X3X3X2Pro(X2X2Pro)nX2(X5)m; where m = 1-5; n = 1-2; o = 2-5; r
CC = 0-8; u = 0-1; X1 = Ile, Leu, Val, Phe, Tyr, Trp or Met; X2 = Trp or
CC Phe; X3 = Arg or Lys; X4 = Trp or Lys; and X5 = Phe, Trp, Arg, Lys or
CC Pro. The peptides are preferably amidated or carboxymethylated. The
CC peptides may be used in methods for inhibiting the growth of a bacterium
CC or yeast, or for inhibiting an endotoxaemia or sepsis associated
CC disorder in a subject. The peptides have a broad activity against
CC antibiotic resistant bacteria, combined with activity against the
CC medically important fungus Candida albicans. In addition, the peptides
CC are useful as antitumour agents and/or antiviral agents. The peptides
CC may be used as sterilants or preservatives of materials susceptible to
CC microbial or viral contamination, e.g. in processed foods to inhibit
CC Salmonella, Yersina and Shigella. The peptides are compact and tend to
CC have a unique polyproline type II extended helix structure that permits
CC them to span the membrane with relatively few amino acids. The peptides
CC possess the ability to work synergistically with antibiotics, and in
CC addition, some of then possess anti-endotoxin activity.
CC N.B. The present sequence represents SEQ ID NO:1 in the claims and
CC examples of the specification, but differs slightly from the SEQ ID NO:1
CC in the sequence listing on page 51 of the specification (see AAW27179).
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 18; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ilkkwpwwpwrrk 13
```

RESULT 2
AAY24609
ID AAY24609 standard; peptide; 13 AA.
XX
AC AAY24609;
XX
DT 18-AUG-1999 (first entry)
XX
DE Indolicidin analogue #61.
XX
KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW additive; shampoo; soap; insecticide; herbicide; preservative;
KW food; technical material.
XX
OS Synthetic.
XX
PN WO9807745-A2.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-US14779.
XX
PR 13-JAN-1997; 97US-0034949.
PR 21-AUG-1996; 96US-0024754.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Erfle D, Fraser JR, Krieger TJ, Taylor R, West MH;
XX
DR WPI; 1998-169090/15.
XX
PT New indolicidin analogues with antimicrobial activity and related
PT nucleic acid - vectors, transformed cells and antibodies, also
PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT toxicity, useful therapeutically, as disinfectants etc.
XX
PS Example 1; Page 32; 129pp; English.
XX
CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC (II), BBBXZXXZXB (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC (V), LBBnXZnXXZnXRK (VI), LKnXZXXZXRRK (VII) and BBXZXXZXBBB (VIII).
CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC infections caused by bacteria (Gram positive or negative, or anaerobic);
CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC trematodes) or viruses. Typical of very many pathogens that can be
CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC derived from the analogues may be used similarly; the compounds may
CC also be prepared from antibiotics or antiarrythmic agents. The analogues
CC may be used therapeutically or to coat medical devices; also they are
CC useful as surface disinfectants, as additives to shampoo or soaps, as
CC insecticides or herbicides, or as preservatives for foods and technical
CC materials. The analogues are administered by injection, lavage, orally
CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC spectrum of activity than indolicidin and modification as compounds
CC reduces their toxicity.
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 19; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ilkkwpwwpwrrk 13
```

RESULT 3
AAW66378
ID AAW66378 standard; peptide; 13 AA.
XX
AC AAW66378;
XX
DT 12-JAN-1999 (first entry)
XX
DE Cationic peptide of claim 15 #5.
XX
KW Indolicidin analogue; resistance; cationic peptide; antibiotic;
KW bacterial infection; tolerance; antibacterial; microorganism;
KW bacteria; fungus; parasite; virus.
XX
OS Synthetic.
XX
PN WO9840401-A2.
XX
PD 17-SEP-1998.
XX
PF 10-MAR-1998; 98WO-CA00190.
XX
PR 25-FEB-1998; 98US-0030619.
PR 10-MAR-1997; 97US-0040649.
PR 20-AUG-1997; 97US-0915314.
PR 26-SEP-1997; 97US-0060099.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Fraser JR, McNicol PJ, West MHP;
XX

DR WPI; 1998-520800/44.
XX
PT New indolicidin peptide analogues - useful for, e.g. enhancing
PT activity of antibiotic or overcoming tolerance, acquired resistance
PT or inherent resistance of microorganisms
XX
PS Claim 15; Page 93; 105pp; English.
XX
CC The present sequence represents a specifically claimed cationic peptide
CC from the present invention. The present invention describes compositions
CC and methods for treating infection, especially bacterial infections. The
CC compositions and methods use cationic peptides in combination with an
CC antibiotic agent which are then administered to a patient to enhance the
CC activity of the antibiotic agent, to overcome: (a) tolerance; (b)
CC acquired resistance; and (c) inherent resistance. The combinations of
CC antibiotics and cationic peptides can provide synergistic activity
CC against a microorganism that is tolerant, inherently resistant, or has
CC acquired resistance to an antibiotic agent. They can be used for killing
CC e.g. bacteria, fungi, parasites and viruses.
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 19; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 4
AAW71690
ID AAW71690 standard; Peptide; 13 AA.
XX
AC AAW71690;
XX
DT 11-JAN-1999 (first entry)
XX
DE Cationic peptide MBI11 (MW 1879).
XX
KW MBI11; cationic peptide; plasmid pKL1; small cryptic plasmid;
KW replication; RepA; vector; RAMP.
XX
OS Synthetic.
XX
PN WO9841636-A2.
XX
PD 24-SEP-1998.
XX
PF 16-MAR-1998; 98WO-CA00214.
XX
PR 14-MAR-1997; 97US-0040722.
XX
PA (BURI/) BURIAN J.
PA (KAYW/) KAY W W.
XX
PI Burian J, Kay WW;
XX
DR WPI; 1998-531571/45.
XX
PT Increasing plasmid copy number in a cell with the repA gene product
PT - and an small cryptic plasmid ori sequence, useful for high level
PT expression of e.g. cytokines, antigens or therapeutic proteins
XX
PS Example 13; Page 54; 82pp; English.
XX
CC MBI11 is a small (mol.wt. 1879) cationic peptide. DNA encoding
CC MBI11 has been incorporated into vector pR2h-B1, in which the
CC replication leader (R21) sequence of RepA (see also AAW71686) is
CC joined to 2 Hpro peptides (see also AAW71692), to provide a
CC vector for expression of MBI11 in host cells. The invention
CC provides controlled replication plasmid vectors (RAMP vectors)
CC comprising a replication origin of a small cryptic plasmid and a
CC gene encoding RepA. The vectors can reach very high levels of
CC plasmid replication, but are not lethal to the host cell, and can
CC be used to direct the high level expression of e.g. cytokines,
CC antigens and therapeutic proteins.
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 19; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 5
AAY94495
ID AAY94495 standard; Peptide; 13 AA.
XX
AC AAY94495;
XX
DT 20-SEP-2000 (first entry)
XX
DE MBI-11 peptide derived from indolicidin.
XX
KW Cellulose binding domain; CBD; cationic peptide;
KW MBI-11; indolicidin; bovine.
XX
OS Bos taurus.
XX
PN WO200031279-A2.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-CA01107.
XX
PR 20-NOV-1998; 98US-0109218.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Burian J, Bartfeld D;
XX
DR WPI; 2000-400086/34.
XX
PT Multi-domain fusion protein expression cassette used for high yield
PT stable production of foreign peptide gene products -
XX
PS Disclosure; Page 24; 73pp; English.
XX
CC A novel method allows the efficient production of cationic peptides in
CC recombinant host cells. The method involves construction of a
CC multi-domain fusion protein expression cassette comprising a promoter and
CC a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC of anionic peptide sequences in the linker sequences neutralises the
CC positive charge of the cationic peptide so that the charge of the
CC fusion protein is controlled. This cassette allows high yield, stable
CC production of the cationic peptide. Cationic peptides such as
CC bovine indolicidin may be used as antimicrobial agents. The present
CC sequence is the MBI-11 peptide. MBI-11 is a cationic peptide derived
CC from modifications of indolicidin.
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 21; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13

```
          ||||||||||||||
Db      1 ilkkwpwwpwrrk 13


RESULT   6
AAY92795
ID   AAY92795 standard; peptide; 13 AA.
XX
AC   AAY92795;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Indolicidin analogue, CP-11.
XX
KW   Magainin; antimicrobial; transgenic plant; protease degradation; Rev4;
KW   indolicidin; protein production; reverse peptide.
XX
OS   Synthetic.
XX
PN   WO200026344-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;   99WO-US25561.
XX
PR   30-OCT-1998;   98US-0106373.
PR   02-NOV-1998;   98US-0106537.
XX
PA   (INTE-) INTERLINK BIOTECHNOLOGIES LLC.
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Everett NP,  Li Q,  Lawrence C,  Davies MH;
XX
DR   WPI; 2000-365597/31.
XX
PT   Polypeptides for reducing proteolytic degradation of proteins
PT   administered to, or produced by a plant comprise indolicin or its
PT   functional equivalents
XX
PS   Disclosure; Page 4; 50pp; English.
XX
CC   Indolicidin is a potent antimicrobial tridecapeptide, originally
CC   purified from cytoplasmic granules of bovine neutrophils. CP-11 is an
CC   analogue, which has better activity against E. coli, Pseudomonas
CC   aeruginosa and Candida albicans, but reduced activity against
CC   Staphylococcus aureus. A reverse peptide, Rev4 (AAY92796) of indolicidin
CC   was found to have increased stability against plant protease degradation.
CC   Expression of antimicrobial peptides in transgenic plants suffers a major
CC   limitation in that the foreign peptides are susceptible to rapid
CC   degradation by proteases. The invention concerns reducing the extent of
CC   protease degradation of a protein applied to, or produced by a plant by
CC   administering indolicidin, Rev4 or a functional equivalent to the plant.
CC   Transgenic plants expressing indolicidin and Rev4 are useful for
CC   production of the antimicrobial peptides. Compositions containing
CC   indolicidin and Rev4 are also useful for production of agronomically
CC   important proteins in plants.
XX
SQ   Sequence   13 AA;


  Query Match             100.0%;  Score 91;  DB 21;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ilkkwpwwpwrrk 13


RESULT   7
AAY91773
ID   AAY91773 standard; Peptide; 13 AA.
XX
AC   AAY91773;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Friedland HD,  Krieger TJ,  Taylor R,  Erfle D,  Fraser JR,  West MHP;
XX
DR   WPI; 2000-223549/19.
XX
PT   Novel pharmaceutical composition containing optionally activated
PT   polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT   -
XX
PS   Disclosure; Page 14; 94pp; English.
XX
CC   This sequence represents a cationic peptide amino acid sequence, which
CC   can be used in the pharmaceutical composition of the invention. The
CC   invention relates to a pharmaceutical composition containing at least one
CC   activated polyoxyalkylene (APO)-modified cationic peptide. The
CC   modification of peptides with APO increases their activity against tumour
CC   cells, including those with a multidrug resistant phenotype. The
CC   pharmaceutical composition can be used to treat tumours, specifically
CC   lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC   cervix, uterus, skin, prostate, liver and colon.
XX
SQ   Sequence   13 AA;


  Query Match             100.0%;  Score 91;  DB 21;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ilkkwpwwpwrrk 13


RESULT   8
AAY91774
ID   AAY91774 standard; Peptide; 13 AA.
XX
AC   AAY91774;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11CN.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
```

```
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR   WPI; 2000-223549/19.
XX
PT   Novel pharmaceutical composition containing optionally activated
PT   polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT   -
XX
PS   Example 3; Page 14; 94pp; English.
XX
CC   This sequence represents a cationic peptide amino acid sequence, which
CC   can be used in the pharmaceutical composition of the invention. The
CC   invention relates to a pharmaceutical composition containing at least one
CC   activated polyoxyalkylene (APO)-modified cationic peptide. The
CC   modification of peptides with APO increases their activity against tumour
CC   cells, including those with a multidrug resistant phenotype. The
CC   pharmaceutical composition can be used to treat tumours, specifically
CC   lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC   cervix, uterus, skin, prostate, liver and colon.
XX
SQ   Sequence   13 AA;


  Query Match           100.0%;  Score 91;  DB 21;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 7.2e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ilkkwpwwpwrrk 13


RESULT  9
AAY91818
ID   AAY91818 standard; Peptide; 13 AA.
XX
AC   AAY91818;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11E1CN.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR   WPI; 2000-223549/19.
XX
PT   Novel pharmaceutical composition containing optionally activated
PT   polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT   -
XX
PS   Disclosure; Page 15; 94pp; English.
XX
CC   This sequence represents a cationic peptide amino acid sequence, which
CC   can be used in the pharmaceutical composition of the invention. The
CC   invention relates to a pharmaceutical composition containing at least one
CC   activated polyoxyalkylene (APO)-modified cationic peptide. The
CC   modification of peptides with APO increases their activity against tumour
CC   cells, including those with a multidrug resistant phenotype. The
CC   pharmaceutical composition can be used to treat tumours, specifically
CC   lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC   cervix, uterus, skin, prostate, liver and colon.
XX
SQ   Sequence   13 AA;


  Query Match           100.0%;  Score 91;  DB 21;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 7.2e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ilkkwpwwpwrrk 13


RESULT  10
AAY91819
ID   AAY91819 standard; Peptide; 13 AA.
XX
AC   AAY91819;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Amino acid sequence of cationic peptide MBI 11E2CN.
XX
KW   Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW   leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW   breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW   multidrug resistance.
XX
OS   Synthetic.
XX
PN   WO9965506-A2.
XX
PD   23-DEC-1999.
XX
PF   14-JUN-1999;   99WO-CA00552.
XX
PR   12-JUN-1998;   98US-0096541.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR   WPI; 2000-223549/19.
XX
PT   Novel pharmaceutical composition containing optionally activated
PT   polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT   -
XX
PS   Disclosure; Page 15; 94pp; English.
XX
CC   This sequence represents a cationic peptide amino acid sequence, which
CC   can be used in the pharmaceutical composition of the invention. The
CC   invention relates to a pharmaceutical composition containing at least one
CC   activated polyoxyalkylene (APO)-modified cationic peptide. The
CC   modification of peptides with APO increases their activity against tumour
CC   cells, including those with a multidrug resistant phenotype. The
CC   pharmaceutical composition can be used to treat tumours, specifically
CC   lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC   cervix, uterus, skin, prostate, liver and colon.
```

XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 21; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 11
AAY91820
ID AAY91820 standard; Peptide; 13 AA.
XX
AC AAY91820;
XX
DT 06-JUN-2000 (first entry)
XX
DE Amino acid sequence of cationic peptide MBI 11E3CN.
XX
KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW multidrug resistance.
XX
OS Synthetic.
XX
PN WO9965506-A2.
XX
PD 23-DEC-1999.
XX
PF 14-JUN-1999; 99WO-CA00552.
XX
PR 12-JUN-1998; 98US-0096541.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Claim 1; Page 15; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 91; DB 21; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 12
AAY24583
ID AAY24583 standard; peptide; 14 AA.
XX
AC AAY24583;
XX
DT 18-AUG-1999 (first entry)
XX
DE Indolicidin analogue #35.
XX
KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW additive; shampoo; soap; insecticide; herbicide; preservative;
KW food; technical material.
XX
OS Synthetic.
XX
PN WO9807745-A2.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-US14779.
XX
PR 13-JAN-1997; 97US-0034949.
PR 21-AUG-1996; 96US-0024754.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Erfle D, Fraser JR, Krieger TJ, Taylor R, West MH;
XX
DR WPI; 1998-169090/15.
XX
PT New indolicidin analogues with antimicrobial activity and related
PT nucleic acid - vectors, transformed cells and antibodies, also
PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT toxicity, useful therapeutically, as disinfectants etc.
XX
PS Claim 13; Page 89; 129pp; English.
XX
CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC (II), BBBXZXXZXB (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC (V), LBBnXZnXXZnXRK (VI), LKnXZXXZXRRK (VII) and BBXZXXZXBBB (VIII).
CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC infections caused by bacteria (Gram positive or negative, or anaerobic);
CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC trematodes) or viruses. Typical of very many pathogens that can be
CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC derived from the analogues may be used similarly; the compounds may
CC also be prepared from antibiotics or antiarrythmic agents. The analogues
CC may be used therapeutically or to coat medical devices; also they are
CC useful as surface disinfectants, as additives to shampoo or soaps, as
CC insecticides or herbicides, or as preservatives for foods and technical
CC materials. The analogues are administered by injection, lavage, orally
CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC spectrum of activity than indolicidin and modification as compounds
CC reduces their toxicity.
XX
SQ Sequence 14 AA;

Query Match 100.0%; Score 91; DB 19; Length 14;
Best Local Similarity 100.0%; Pred. No. 7.8e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 13
AAY91811
ID AAY91811 standard; Peptide; 14 AA.
XX
AC AAY91811;
XX
DT 06-JUN-2000 (first entry)
XX
DE Amino acid sequence of cationic peptide MBI 11D11H.
XX
KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KW multidrug resistance.
XX
OS Synthetic.
XX
PN WO9965506-A2.
XX
PD 23-DEC-1999.
XX
PF 14-JUN-1999; 99WO-CA00552.
XX
PR 12-JUN-1998; 98US-0096541.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Disclosure; Page 15; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX
SQ Sequence 14 AA;

Query Match 100.0%; Score 91; DB 21; Length 14;
Best Local Similarity 100.0%; Pred. No. 7.8e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 14
AAY24582
ID AAY24582 standard; peptide; 21 AA.
XX
AC AAY24582;
XX
DT 18-AUG-1999 (first entry)
XX
DE Indolicidin analogue #34.
XX
KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
KW antimicrobial; antibiotic; antiarrythmic; surface disinfectant;
KW additive; shampoo; soap; insecticide; herbicide; preservative;
KW food; technical material.
XX
OS Synthetic.
XX
PN WO9807745-A2.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-US14779.
XX
PR 13-JAN-1997; 97US-0034949.
PR 21-AUG-1996; 96US-0024754.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Erfle D, Fraser JR, Krieger TJ; Taylor R, West MH;
XX
DR WPI; 1998-169090/15.
XX
PT New indolicidin analogues with antimicrobial activity and related
PT nucleic acid - vectors, transformed cells and antibodies, also
PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
PT toxicity, useful therapeutically, as disinfectants etc.
XX
PS Claim 13; Page 89; 129pp; English.
XX
CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
CC (I)-(VIII) containing up to 25 amino acids (aa): RXZXXZXB (I), BXZXXZXB
CC (II), BBBXZXXZXB (III), BXZXXZXBBBn(AA)nMILBBAGS (IV), BXZXXZXBB(AA)nM
CC (V), LBBnXZnXXZnXRK (VI), LKnXZXXZXRRK (VII) and BBXZXXZXBBB (VIII).
CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
CC infections caused by bacteria (Gram positive or negative, or anaerobic);
CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
CC trematodes) or viruses. Typical of very many pathogens that can be
CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
CC derived from the analogues may be used similarly; the compounds may
CC also be prepared from antibiotics or antiarrythmic agents. The analogues
CC may be used therapeutically or to coat medical devices; also they are
CC useful as surface disinfectants, as additives to shampoo or soaps, as
CC insecticides or herbicides, or as preservatives for foods and technical
CC materials. The analogues are administered by injection, lavage, orally
CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
CC spectrum of activity than indolicidin and modification as compounds
CC reduces their toxicity.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 91; DB 19; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRRK 13
||||||||||||||
Db 1 ilkkwpwwpwrrk 13

RESULT 15
AAY91806
ID AAY91806 standard; Peptide; 21 AA.
XX
AC AAY91806;
XX
DT 06-JUN-2000 (first entry)
XX
DE Amino acid sequence of cationic peptide MBI 11D4CN.
XX
KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;

KW multidrug resistance.
XX
OS Synthetic.
XX
PN WO9965506-A2.
XX
PD 23-DEC-1999.
XX
PF 14-JUN-1999; 99WO-CA00552.
XX
PR 12-JUN-1998; 98US-0096541.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;
XX
DR WPI; 2000-223549/19.
XX
PT Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
PT -
XX
PS Disclosure; Page 15; 94pp; English.
XX
CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 91; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ILKKWPWWPWRRK 13
           |||||||||||||
Db       1 ilkkwpwwpwrrk 13
```

Search completed: January 4, 2002, 08:40:26
Job time: 109 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:40:57 ; Search time 24.75 Seconds
(without alignments)
11.820 Million cell updates/sec

Title: US-09-444-281-35
Perfect score: 91
Sequence: 1 ILKKWPWWPWRRK 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	91	100.0	13	4	US-08-915-314-30	Sequence 30, Appl
2	91	100.0	13	4	US-08-915-314-62	Sequence 62, Appl
3	91	100.0	13	4	US-08-915-314-63	Sequence 63, Appl
4	91	100.0	13	4	US-08-915-314-64	Sequence 64, Appl
5	91	100.0	13	4	US-09-042-071-36	Sequence 36, Appl
6	91	100.0	14	4	US-08-915-314-57	Sequence 57, Appl
7	91	100.0	21	4	US-08-915-314-54	Sequence 54, Appl
8	87	95.6	12	4	US-08-915-314-52	Sequence 52, Appl
9	86	94.5	12	4	US-08-915-314-74	Sequence 74, Appl
10	86	94.5	12	4	US-08-702-054B-5	Sequence 5, Appli
11	86	94.5	13	4	US-08-915-314-58	Sequence 58, Appl
12	86	94.5	14	4	US-08-915-314-59	Sequence 59, Appl
13	86	94.5	15	4	US-08-702-054B-40	Sequence 40, Appl
14	86	94.5	20	4	US-08-915-314-55	Sequence 55, Appl
15	86	94.5	21	4	US-08-915-314-56	Sequence 56, Appl
16	85	93.4	12	4	US-08-915-314-69	Sequence 69, Appl
17	85	93.4	13	4	US-08-915-314-38	Sequence 38, Appl
18	85	93.4	13	4	US-08-915-314-45	Sequence 45, Appl
19	85	93.4	13	4	US-08-702-054B-1	Sequence 1, Appli
20	85	93.4	13	4	US-08-702-054B-17	Sequence 17, Appl
21	85	93.4	13	4	US-08-702-054B-32	Sequence 32, Appl
22	83	91.2	12	4	US-08-915-314-24	Sequence 24, Appl
23	83	91.2	13	4	US-08-915-314-49	Sequence 49, Appl
24	83	91.2	13	4	US-08-915-314-50	Sequence 50, Appl
25	83	91.2	13	4	US-08-915-314-51	Sequence 51, Appl
26	83	91.2	13	4	US-08-702-054B-30	Sequence 30, Appl
27	83	91.2	13	4	US-08-702-054B-31	Sequence 31, Appl
28	83	91.2	13	4	US-08-702-054B-34	Sequence 34, Appl
29	83	91.2	13	4	US-08-702-054B-35	Sequence 35, Appl
30	82	90.1	13	4	US-08-915-314-25	Sequence 25, Appl
31	82	90.1	13	4	US-08-915-314-66	Sequence 66, Appl
32	82	90.1	13	4	US-08-915-314-67	Sequence 67, Appl
33	82	90.1	13	4	US-08-702-054B-33	Sequence 33, Appl
34	81	89.0	11	4	US-08-915-314-75	Sequence 75, Appl
35	81	89.0	15	4	US-08-702-054B-39	Sequence 39, Appl
36	80	87.9	14	4	US-08-702-054B-18	Sequence 18, Appl
37	80	87.9	15	4	US-08-702-054B-41	Sequence 41, Appl
38	80	87.9	16	4	US-08-702-054B-2	Sequence 2, Appli
39	79.5	87.4	16	4	US-08-702-054B-38	Sequence 38, Appl
40	79	86.8	17	4	US-08-702-054B-42	Sequence 42, Appl
41	78	85.7	11	4	US-08-915-314-28	Sequence 28, Appl
42	78	85.7	12	4	US-08-915-314-40	Sequence 40, Appl
43	77	84.6	12	4	US-08-915-314-39	Sequence 39, Appl
44	77	84.6	12	4	US-08-702-054B-27	Sequence 27, Appl
45	76	83.5	12	4	US-08-915-314-77	Sequence 77, Appl

ALIGNMENTS

```
RESULT 1
US-08-915-314-30
; Sequence 30, Application US/08915314
; Patent No. 6180604
; GENERAL INFORMATION:
;   APPLICANT: Fraser, Janet R.
;   APPLICANT: West, Michael H.P.
;   APPLICANT: Krieger, Timothy J.
;   APPLICANT: Taylor, Robert
;   APPLICANT: Erfle, Douglas
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
;   TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
;   NUMBER OF SEQUENCES: 90
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SEED and BERRY LLP
;     STREET: 6300 Columbia Center, 701 Fifth Avenue
;     CITY: Seattle
;     STATE: Washington
;     COUNTRY: USA
;     ZIP: 98104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/915,314
;     FILING DATE: 20-AUG-1997
;     CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;     NAME: No. 6180604tenburg Ph.D., Carol
;     REGISTRATION NUMBER: 39,317
;     REFERENCE/DOCKET NUMBER: 660081.405
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (206) 622-4900
;     TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 30:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 13 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
US-08-915-314-30


  Query Match            100.0%;  Score 91;  DB 4;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 5.4e-07;
  Matches  13;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 ILKKWPWWPWRRK 13
```

```
              ||||||||||||
Db      1 ILKKWPWWPWRRK 13


RESULT   2
US-08-915-314-62
; Sequence 62, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  62:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1
;      OTHER INFORMATION:  /note= "D-Form of Isoleucine"
US-08-915-314-62


  Query Match             100.0%;  Score 91;  DB 4;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ILKKWPWWPWRRK 13


RESULT   3
US-08-915-314-63
; Sequence 63, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  13
;      OTHER INFORMATION:  /note= "D-Form of Lysine"
US-08-915-314-63


  Query Match             100.0%;  Score 91;  DB 4;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWPWRRK 13
          |||||||||||||
Db      1 ILKKWPWWPWRRK 13


RESULT   4
US-08-915-314-64
; Sequence 64, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/915,314
;       FILING DATE:  20-AUG-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6180604tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  660081.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  64:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  13 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  1
;       OTHER INFORMATION:  /note= "D-Form of Isoleucine"
;     FEATURE:
;       NAME/KEY:  Modified-site
;       LOCATION:  13
;       OTHER INFORMATION:  /note= "D-Form of Lysine"
US-08-915-314-64


  Query Match             100.0%;  Score 91;  DB 4;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRRK 13
           |||||||||||||
Db       1 ILKKWPWWPWRRK 13


RESULT   5
US-09-042-071-36
; Sequence 36, Application US/09042071
; Patent No. 6294372
;  GENERAL INFORMATION:
;    APPLICANT:  Burian, Jan
;    APPLICANT:  Kay, William W.
;    TITLE OF INVENTION:  REPLICATION GENES AND GENE PRODUCTS FROM
;    TITLE OF INVENTION:  SMALL CRYPTIC PLASMIDS AND METHODS FOR CONSTRUCTING
;    TITLE OF INVENTION:  CONTROLLED-REPLICATION PLASMID VECTORS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,071
;      FILING DATE:  13-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  660081.407
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-042-071-36


  Query Match             100.0%;  Score 91;  DB 4;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRRK 13
           |||||||||||||
Db       1 ILKKWPWWPWRRK 13


RESULT   6
US-08-915-314-57
; Sequence 57, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-57


  Query Match             100.0%;  Score 91;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRRK 13
           |||||||||||||
Db       1 ILKKWPWWPWRRK 13


RESULT   7
US-08-915-314-54
; Sequence 54, Application US/08915314
; Patent No. 6180604
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-54


  Query Match             100.0%;  Score 91;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-07;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRRK 13
           |||||||||||||
Db       1 ILKKWPWWPWRRK 13


RESULT   8
US-08-915-314-52
; Sequence 52, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-52


  Query Match             95.6%;  Score 87;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LKKWPWWPWRRK 13
           ||||||||||||
Db       1 LKKWPWWPWRRK 12


RESULT   9
US-08-915-314-74
; Sequence 74, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  74:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
```

US-08-915-314-74

Query Match 94.5%; Score 86; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRR 12
||||||||||||
Db 1 ILKKWPWWPWRR 12

RESULT 10
US-08-702-054B-5
; Sequence 5, Application US/08702054B
; Patent No. 6191254
; GENERAL INFORMATION:
; APPLICANT: Falls, Timothy J.
; APPLICANT: Hancock, Robert E. W.
; APPLICANT: Gough, Monisha
; TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES
; TITLE OF INVENTION: AND METHODS OF SCREENING FOR THE SAME
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,054B
; FILING DATE: 23-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/002,687
; FILING DATE: 23-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07420/013001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-702-054B-5

Query Match 94.5%; Score 86; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRR 12
||||||||||||
Db 1 ILKKWPWWPWRR 12

RESULT 11
US-08-915-314-58
; Sequence 58, Application US/08915314
; Patent No. 6180604
; GENERAL INFORMATION:
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfle, Douglas
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
; NUMBER OF SEQUENCES: 90
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,314
; FILING DATE: 20-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6180604tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 660081.405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-915-314-58

Query Match 94.5%; Score 86; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILKKWPWWPWRR 12
||||||||||||
Db 1 ILKKWPWWPWRR 12

RESULT 12
US-08-915-314-59
; Sequence 59, Application US/08915314
; Patent No. 6180604
; GENERAL INFORMATION:
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfle, Douglas
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
; NUMBER OF SEQUENCES: 90
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```
;       APPLICATION NUMBER:  US/08/915,314
;       FILING DATE:  20-AUG-1997
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6180604tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  660081.405
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  59:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  14 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-915-314-59


  Query Match             94.5%;  Score 86;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRR 12
           ||||||||||||
Db       1 ILKKWPWWPWRR 12


RESULT  13
US-08-702-054B-40
; Sequence 40, Application US/08702054B
; Patent No. 6191254
;  GENERAL INFORMATION:
;    APPLICANT:  Falls, Timothy J.
;    APPLICANT:  Hancock, Robert E. W.
;    APPLICANT:  Gough, Monisha
;    TITLE OF INVENTION:  ANTIMICROBIAL CATIONIC PEPTIDES
;    TITLE OF INVENTION:  AND METHODS OF SCREENING FOR THE SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/702,054B
;      FILING DATE:  23-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/002,687
;      FILING DATE:  23-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07420/013001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-702-054B-40


  Query Match             94.5%;  Score 86;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 3e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRR 12
           ||||||||||||
Db       1 ILKKWPWWPWRR 12


RESULT  14
US-08-915-314-55
; Sequence 55, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-55


  Query Match             94.5%;  Score 86;  DB 4;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRR 12
           ||||||||||||
Db       1 ILKKWPWWPWRR 12


RESULT  15
US-08-915-314-56
; Sequence 56, Application US/08915314
; Patent No. 6180604
;  GENERAL INFORMATION:
;    APPLICANT:  Fraser, Janet R.
;    APPLICANT:  West, Michael H.P.
;    APPLICANT:  Krieger, Timothy J.
;    APPLICANT:  Taylor, Robert
;    APPLICANT:  Erfle, Douglas
```

```
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATING
;    TITLE OF INVENTION:  INFECTIONS USING ANALOGUES OF INDOLICIDIN
;    NUMBER OF SEQUENCES:  90
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,314
;      FILING DATE:  20-AUG-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6180604tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  660081.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  21 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-915-314-56


  Query Match             94.5%;  Score 86;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-06;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWRR 12
           ||||||||||||
Db       1 ILKKWPWWPWRR 12


Search completed: January  4, 2002, 08:40:57
Job time: 140 sec
```

```
OM of: US-09-444-281-35 to: GenEmbl:*   out_format : pfs

Date: Jan 4, 2002 10:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084143_16195/app_query.fasta_1.210
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09444281_@CGN1_1_5145 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-35
Query length: 13
Database: GenEmbl:*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1728.450000


score_list:
Sequence                Strd Orig   ZScore   EScore Len.  ! Documentation  ..
gb_om:BTINDLCD            +   73.00  158.82  1.43      550  ! X67340 B.taurus mRNA for indoli
gb_pat:AX098418           +   73.00  141.25  13.65     6446  ! AX098418 Sequence 5 from Paten
gb_htg:AP003754           -   68.00  109.17  836.02    129052  ! AP003754 Oryza sativa chromo
gb_ro:MMU08210            +   66.00  132.63  41.24     2651  ! U08210 Mus musculus tropoelast
gb_ro:AF289665            -   66.00  106.21  1.2e+03   107257  ! AF289665 Mus musculus EIF4H
gb_htg:AC091250           +   66.00  101.74  2.2e+03   200849  ! AC091250 Mus musculus chromo
gb_htg:AC020365           +   65.00  113.72  466.29    27780  ! AC020365 Drosophila melanogas
gb_in:AC008316            +   65.00  101.18  2.3e+03   160817  ! AC008316 Drosophila melanoga
gb_in:AC008315            +   65.00  100.50  2.5e+03   177028  ! AC008315 Drosophila melanoga
gb_in:AE003684            +   65.00   98.96  3.1e+03   219579  ! AE003684 Drosophila melanoga
gb_ro:MMA2IXCOA           -   64.00  114.12  443.17    19479  ! Z22923 M.musculus alpha2 (IX)
gb_pr:AL162423            -   64.00  103.33  1.8e+03   88323  ! AL162423 Human DNA sequence f
gb_htg:AL356097_0         -   64.00  101.76  2.2e+03   110000  ! AL356097 Homo sapiens chromo
gb_htg:AC027113           +   64.00   99.11  3.0e+03   159391  ! AC027113 Homo sapiens chromo
gb_htg:AL358473           +   64.00   97.94  3.5e+03   187845  ! AL358473 Homo sapiens chromo
gb_ro:AF210429            +   63.00  138.78  18.73     456  ! AF210429 Mus musculus group X s
gb_ro:AF166097            +   63.00  132.90  39.85     1040  ! AF166097 Mus musculus group X
gb_ro:AF191309            -   63.00  124.85  111.95    3213  ! AF191309 Mus musculus zinc fin
gb_ro:RNY17153            +   63.00  120.98  183.83    5523  ! Y17153 Rattus norvegicus mRNA
gb_ro:RATNCHVS            +   63.00  119.33  227.10    6957  ! M26643 Rat skeletal muscle vol
gb_in:CEF46A8             +   63.00  112.95  515.06    17015  ! Z81539 Caenorhabditis elegans
gb_ba:AE004649            +   63.00  110.66  690.23    23425  ! AE004649 Pseudomonas aeruginc
gb_ro:AC003063            -   63.00   98.53  3.3e+03   128143  ! AC003063 Mus musculus Chromo
gb_ro:AC005817            -   63.00   95.77  4.7e+03   188707  ! AC005817 Mus musculus chromo
gb_htg:AC078895           -   63.00   94.90  5.2e+03   213204  ! AC078895 Mus musculus chromo
gb_htg:AC079043           +   63.00   94.85  5.2e+03   214723  ! AC079043 Mus musculus chromo
gb_pat:AX078941           +   62.00  130.84  51.92     1029  ! AX078941 Sequence 17 from Pate
gb_pl:ATPHYB              +   61.50  120.35  199.31    3850  ! X17342 Arabidopsis thaliana ph
gb_pl:ATHPHTOCHB          +   61.50  116.60  322.37    6509  ! L09262 Arabidopsis thaliana ph
gb_in:CELM01G5            +   61.50  104.06  1.6e+03   37675  ! AF078786 Caenorhabditis elega
gb_pl:AC005724            +   61.50   98.12  3.5e+03   86671  ! AC005724 Arabidopsis thaliana
gb_pl:AP003141            -   61.50   94.63  5.4e+03   141275  ! AP003141 Oryza sativa genomi
gb_htg:AC013145           +   61.00  104.69  1.5e+03   29692  ! AC013145 Drosophila melanogas
gb_in:DMBR37P7            -   61.00   95.84  4.6e+03   102619  ! AL050231 Drosophila melanoga
gb_htg:AC010475           -   61.00   94.44  5.5e+03   124927  ! AC010475 Homo sapiens chromo
gb_htg:AC092745           -   61.00   94.43  5.5e+03   125047  ! AC092745 Homo sapiens chromo
gb_pr:AC010485            -   61.00   94.41  5.6e+03   125439  ! AC010485 Homo sapiens chromo
gb_htg:AC080046           -   61.00   93.54  6.2e+03   141674  ! AC080046 Homo sapiens chromo
gb_htg:AP003910           +   61.00   93.44  6.3e+03   143549  ! AP003910 Oryza sativa chromo
gb_htg:AC083933           +   61.00   93.44  6.3e+03   143717  ! AC083933 Homo sapiens chromo
gb_pl:AP001366            +   61.00   93.32  6.4e+03   146081  ! AP001366 Oryza sativa genomi
gb_htg:AC068010           -   61.00   92.42  7.2e+03   165652  ! AC068010 Homo sapiens clo
gb_htg:AC063927           -   61.00   92.26  7.3e+03   169491  ! AC063927 Homo sapiens chr
gb_htg:AC067797           -   61.00   91.99  7.6e+03   175940  ! AC067797 Homo sapiens chr
gb_in:AE003417            -   61.00   87.84  1.3e+04   314661  ! AE003417 Drosophila melan

seq_name: gb_om:BTINDLCD

seq_documentation_block:
LOCUS       BTINDLCD      550 bp    mRNA            MAM       07-OCT-1992
DEFINITION  B.taurus mRNA for indolicidin.
ACCESSION   X67340
VERSION     X67340.1  GI:462
KEYWORDS    indolicidin.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Del Sal,G., Storici,P., Schneider,C., Romeo,D. and Zanetti,M.
  TITLE     cDNA cloning of the neutrophil bactericidal peptide indolicidin
  JOURNAL   Biochem. Biophys. Res. Commun. 187 (1), 467-472 (1992)
  MEDLINE   92392368
REFERENCE   2  (bases 1 to 550)
  AUTHORS   Del Sal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1992) G. Del Sal, Univ. of Trieste, Dip. di
            Biochimica, Biofisica e Chimica, delle Macromolecole, Via A.
            Valerio, 38, 34127 Trieste, ITALY
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_line="bone marrow cells"
     sig_peptide     13. .99
     CDS             13. .447
                     /codon_start=1
                     /product="indolicin prepropeptide"
                     /protein_id="CAA47755.1"
                     /db_xref="GI:463"
                     /db_xref="SWISS-PROT:P33046"
                     /translation="MQTQRASLSLGRWSLWLLLLGLVVPSASAQALSYREAVLRAVDQ
                     LNELSSEANLYRLLELDPPPKDNEDLGTRKPVSFTVKETVCPRTIQQPAEQCDFKEKG
                     RVKQCVGTVTLDPSNDQFDLNCNELQSVILPWKWPWWPWRRG"
     mat_peptide     403. .444
                     /product="indolicin"
BASE COUNT      126 a    140 c    169 g    115 t
ORIGIN


alignment_scores:
           Quality:   73.00          Length:     9
             Ratio:  8.111            Gaps:     0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x BTINDLCD   ..

 Align seg 1/1  to: BTINDLCD  from: 1  to: 550

       4 LysTrpProTrpTrpProTrpArgArg 12
         |||||||||||||||||||||||||||
     415 AAATGGCCATGGTGGCCTTGGCGCAGA 441

seq_name: gb_pat:AX098418

seq_documentation_block:
LOCUS       AX098418     6446 bp    mRNA            PAT       03-APR-2001
DEFINITION  Sequence 5 from Patent WO0119969.
ACCESSION   AX098418
VERSION     AX098418.1  GI:13537710
KEYWORDS    .
SOURCE      Nicotiana benthamiana.
```

```
  ORGANISM  Nicotiana benthamiana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (bases 1 to 6446)
  AUTHORS   Garger,S.J., Holtz,B.R., Mcculloch,M.J. and Turpen,T.H.
  TITLE     A process for isolating and purifying viruses, soluble proteins and
            peptides from plant sources
  JOURNAL   Patent: WO 0119969-A 5 22-MAR-2001;
            Large Scale Biology Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .6446
                     /organism="Nicotiana benthamiana"
                     /db_xref="taxon:4100"
BASE COUNT     1873 a   1234 c   1563 g   1776 t
ORIGIN


alignment_scores:
            Quality:   73.00            Length:      9
              Ratio:  8.111              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AX098418   ..

 Align seg 1/1  to: AX098418  from: 1  to: 6446

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||
    6213 AAGTGGCCTTGGTGGCCATGGCGCCGA 6239

seq_name: gb_htg:AP003754

seq_documentation_block:
LOCUS       AP003754   129052 bp    DNA             HTG       14-JUN-2001
DEFINITION  Oryza sativa chromosome 7 clone OJ1341_A08, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003754
VERSION     AP003754.1  GI:14422472
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1341_A08.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1341_A08
  JOURNAL   Published Only in DataBase (2001) In press
REFERENCE   2  (bases 1 to 129052)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .129052
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="7"
                     /clone="OJ1341_A08"
BASE COUNT     35088 a   28842 c   28921 g   36151 t      50 others
ORIGIN


alignment_scores:
            Quality:   68.00            Length:     11
              Ratio:  6.800              Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-35 x AP003754/rev   ..

 Align seg 1/1  to reverse of: AP003754  from: 1  to: 129052

       3 LysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::||||||||||||||||||||   |||:::
   53246 CGCCGCTGGCCTTGGTGGCCCTGGACGCGGCGG 53214

seq_name: gb_ro:MMU08210

seq_documentation_block:
LOCUS       MMU08210     2651 bp    mRNA            ROD       31-OCT-1995
DEFINITION  Mus musculus tropoelastin mRNA, complete cds.
ACCESSION   U08210
VERSION     U08210.1  GI:473273
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2651)
  AUTHORS   Wydner,K.S., Sechler,J.L., Boyd,C.D. and Passmore,H.C.
  TITLE     Use of an intron polymorphism to localize the tropoelastin gene to
            mouse chromosome 5 in a region of linkage conservation with human
            chromosome 7
  JOURNAL   Genomics 23 (1), 125-131 (1994)
  MEDLINE   95130069
REFERENCE   2  (bases 1 to 2651)
  AUTHORS   Boyd,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1994) Charles D. Boyd, Department of Surgery,
            UMDNJ - Robert Wood Johnson Medical School, 51 French St., New
            Brunswick, NJ 08903, USA
FEATURES             Location/Qualifiers
     source          1. .2651
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="lung"
                     /dev_stage="adult"
     CDS             1. .2583
                     /codon_start=1
                     /product="tropoelastin"
                     /protein_id="AAA80155.1"
                     /db_xref="GI:473274"
                     /translation="MAGLTAVVPQPGVLLILLLNLLHPAQPGGVPGAVPGGLPGGVPG
                     GVYYPGAGIGGLGGGGGALGPGGKPPKPGAGLLGTFGAGPGGLGGAGPGAGLGAFPAG
                     TFPGAGALVPGGAAGAAAAYKAAAKAGAGLGGVGGVPGGVGVGGVPGGVGVGGVPGGV
                     GVGGVPGGVGGIGGIGGLGVSTGAVVPQVGAGIGAGGKPGKVPGVGLPGVYPGGVLPG
                     TGARFPGVGVLPGVPTGTGVKAKAPGGGGAFSGIPGVGPFGGQQPGVPLGYPIKAPKL
                     PGGYGLPYTNGKLPYGVAGAGGKAGYPTGTGVGSQAAAAAAKAAKYGAGGAGVLPGVG
                     GGGIPGGAGAIPGIGGIAGAGTPAAAAAAKAAAKAAKYGAAGGLVPGGPGVRLPGAGI
                     PGVGGIPGVGGIPGVGGPGIGGPGIVGGPGAVSPAAAAKAAAKAAKYGARGGVGIPTY
                     GVGAGGFPGYGVGAGAGLGGASPAAAAAAAKAAKYGAGGAGALGGLVPGAVPGALPGA
                     VPAVPGAGGVPGAGTPAAAAAAAAAKAAAKAGLGPGVGGVPGGVGVGGIPGGVGVGGV
                     PGGVGPGGVTGIGAGPGGLGGAGSPAAAKSAAKAAAKAQYRAAAGLGAGVPGFGAGAG
                     VPGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGGLGGPG
```

```
                    GLGGPGGLGGPGGLGGAGVPGRVAGAAPPAAAAAAAKAAAKAAQYGLGGAGGLGAGGL
                    GAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGGVSPAAAAKAAKYGAAGLGGVLGA
                    RPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPYGGALGALGYQGGGCFGKSCG
                    RKRK"
BASE COUNT      410 a    623 c   1029 g    589 t
ORIGIN


alignment_scores:
           Quality:   66.00          Length:     10
             Ratio:  7.333            Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-35 x MMU08210   ..

 Align seg 1/1  to: MMU08210  from: 1  to: 2651

       4 LysTrpProTrpTrpProTrpArgArgLys 13
         :::||||||||||||||||||||   :::
    2025 AGGTGGCCTTGGTGGCCCTGGAGGTCTCGG 2054

seq_name: gb_ro:AF289665

seq_documentation_block:
LOCUS       AF289665   107257 bp    DNA             ROD       14-AUG-2000
DEFINITION  Mus musculus EIF4H gene, partial cds; LIMK1 gene, complete cds; and
            ELN gene, partial cds.
ACCESSION   AF289665
VERSION     AF289665.1  GI:9800517
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 107257)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Genome Technology Branch, National Human
            Genome Research Institute, 49 Convent Dr. Rm. 2A02, Bethesda, MD
            20892, USA
FEATURES             Location/Qualifiers
     source          1. .107257
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="42j20"
                     /clone_lib="Research Genetics CITB-CJ7-B BAC library"
     mRNA            complement(<7676. .7763)
                     /product="EIF4H"
     CDS             complement(<7676. .7734)
                     /codon_start=1
                     /product="EIF4H"
                     /protein_id="AAF99335.1"
                     /db_xref="GI:9800519"
                     /translation="MADFDTYDDRAYSSFGGGRG"
     mRNA            complement(join(24480. .25812,25920. .26077,26336. .26391,
                     28274. .28430,30220. .30285,30530. .30589,32612. .32743,
                     33634. .33720,34257. .34440,37341. .37507,37617. .37722,
                     38745. .38951,41098. .41207,41439. .41577,55071. .55167,
                     56768. .57026))
                     /product="LIMK1"
     CDS             complement(join(25650. .25812,25920. .26077,26336. .26391,
                     28274. .28430,30220. .30285,30530. .30589,32612. .32743,
                     33634. .33720,34257. .34440,37341. .37507,37617. .37722,
                     38745. .38951,41098. .41207,41439. .41577,55071. .55167,
                     56768. .56822))
                     /codon_start=1
                     /evidence=not_experimental
                     /product="LIMK1"
                     /protein_id="AAF99334.1"
                     /db_xref="GI:9800518"
                     /translation="MRLTLLCCTWREERMGEEGSELPVCASCGQRIYDGQYLQALNAD
                     WHADCFRCCECSVSLSHQYYEKDGQLFCKKDYWARYGESCHGCSEHITKGLVMVAGEL
                     KYHPECFICLACGNFIGDGDTYTLVEHSKLYCGQCYYQTVVTPVIEQILPDSPGSHLP
                     HTVTLVSIPASAHGKRGLSVSIDPPHGPPGCGTEHSHTVRVQGVDPGCMSPDVKNSIH
                     VGDRILEINGTPIRNVPLDEIDLLIQETSRLLQLTLEHDPHDSLGHGPVSDPSPLSSP
                     VHTPSGQAASSARQKPVLRSCSIDTSPGTSSLASPASQRKDLGRSESLRVVCRPHRIF
                     RPSDLIHGEVLGKGCFGQAIKVTHRETGEVMVMKELIRFDEETQRTFLKEVKVMRCLE
                     HPNVLKFIGVLYKDKRLNFITEYIKGGTLRGIIKNMDSQYPWSQRVSFAKDIASGMAY
                     LHSMNIIHRDLNSHNCLVRENRNVVVADFGLARLMIDEKNQSEDLRSLKKPDRKKRYT
                     VVGNPYWMAPEMINGRSYDEKVDVFSFGIVLCEIIGRVNADPDYLPRTMDFGLNVRGF
                     LDRYCPPNCPPSFFPITVRCCDLDPEKRPSFVKLEQWLETLRMHLSGHLPLGPQLEQL
                     ERGFWETYRRGESSLPAHPEVPD"
     mRNA            complement(join(72144. .72255,72961. .73011,73719. .73748,
                     74172. .74216,74796. .74846,74997. .75035,75941. .76081,
                     76252. .76314,76403. .76504,78538. .78576,79493. .79651,
                     80837. .80896,80981. .81121,82426. .82485,83678. .83779,
                     84115. .84162,84859. .84948,85423. .85476,86452. .86598,
                     87420. .87479,87564. .87653,90048. .90095,90466. .90510,
                     90895. .90936,93350. .93421,94200. .94229,94899. .94970,
                     96821. .96868,97559. .97621,97702. .97863,98076. .98177,
                     99496. .99531,100964. .100993,104734. .104766,
                     105595. .>105630))
                     /note="GenScan; P1=0.902, P2=0.979, P3=0.899, P4=0.797,
                     P5=0.640, P6=0.421, P7=0.897, P8=0.912, P9=0.991,
                     P10=0.999, P11=0.630, P12=0.813, P13=0.935, P14=0.767,
                     P15=0.773, P16=0.567, P17=0.806, P18=0.998, P19=0.999,
                     P20=0.996, P21=0.978, P22=0.980, P23=0.999, P24=0.986,
                     P25=0.579, P26=0.757, P27=0.661, P28=0.985, P29=0.952,
                     P30=0.228, P31=0.203, P32=0.451, P33=0.463, P34=0.525,
                     P35=0.228, P36=0.969, P37=0.995, P38=0.905, P39=0.655,
                     P40=0.986, P41=0.973, P42=0.930, P43=0.967, P44=0.995,
                     P45=0.999, P46=0.351, P47=0.994, P48=0.534, P49=0.787,
                     P50=0.943, P51=PlyA"
                     /product="ELN"
     CDS             complement(join(72212. .72255,72961. .73011,73719. .73748,
                     74172. .74216,74796. .74846,74997. .75035,75941. .76081,
                     76252. .76314,76403. .76504,78538. .78576,79493. .79651,
                     80837. .80896,80981. .81121,82426. .82485,83678. .83779,
                     84115. .84162,84859. .84948,85423. .85476,86452. .86598,
                     87420. .87479,87564. .87653,90048. .90095,90466. .90510,
                     90895. .90936,93350. .93421,94200. .94229,94899. .94970,
                     96821. .96868,97559. .97621,97702. .97863,98076. .98177,
                     99496. .99531,100964. .100993,104734. .104766,
                     105595. .>105630))
                     /codon_start=3
                     /product="ELN"
                     /protein_id="AAF99336.1"
                     /db_xref="GI:9800520"
                     /translation="AGIGGLGGGGGALGPGGKPPKPGAGLLGTFGAGPGGLGGAGPGA
                     GLGAFPAGTFPGAGALVPGGAAGAAAAYKAAAKAGAGLGGVGGVPGGVGVGGVPGGVG
                     VGGVPGGVGVGGVPGGVGGIGGIGGLGVSTGAVVPQVGAGIGAGGKPGKVPGVGLPGV
                     YPGGVLPGTGARFPGVGVLPGVPTGTGVKAKAPGGGGAFAGIPGVGPFGGQQPGVPLG
                     YPIKAPKLPGGYGLPYTNGKLPYGVAGAGGKAGYPTGTGVGSQAAAAAAKAAKYGAGG
                     AGVLPGVGGGGIPGGAGAIPGIGGIAGAGTPAAAAAAKAAAKAAKYGAAGGLVPGGPG
                     VRLPGAGIPGVGGIPGVGGPGIGGPGIVGGPGAVSPAAAAKAAAKAAKYGAR
                     GGVGIPTYGVGAGGFPGYGVGAGAGLGGASPAAAAAAAKAAKYGAGGAGALGGLVPGA
                     VPGALPGAVPAVPGAGGVPGAGTPAAAAAAAAAKAAAKAGLGPGVGGVPGGVGVGGIP
                     GGVGVGGVPGGVGPGGVTGIGAGPGGLGGAGSPAAAKSAAKAAAKAQYRAAAGLGAGV
                     PGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAVPGSLAASKAAKYGA
                     AGGLGGPGGLGGPGGLGGPGGLGGAGVPGRVAGAAPPAAAAAAAKAAAKAAQYGLGGA
                     GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGGVSPAAAAKAAKYGAA
                     GLGGVLGARPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPYGGALGALGYQGG
                     GCFGKSCGRKRK"
BASE COUNT    26300 a   26393 c   26598 g   27966 t
ORIGIN


alignment_scores:
           Quality:   66.00          Length:     10
             Ratio:  7.333            Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
```

```
 US-09-444-281-35 x AF289665/rev  ..

 Align seg 1/1  to reverse of: AF289665  from: 1  to: 107257

       4 LysTrpProTrpTrpProTrpArgArgLys 13
         :::||||||||||||||||||||||   :::
   76497 AGGTGGCCTTGGTGGCCCTGGAGGTCTCGG 76468

seq_name: gb_htg:AC091250

seq_documentation_block:
LOCUS       AC091250   200849 bp    DNA             HTG       11-APR-2001
DEFINITION  Mus musculus chromosome 5 clone RP23-315E2 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION   AC091250
VERSION     AC091250.1  GI:13592171
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200849)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
            Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
            Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200849)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                Center project name: aty
                Center clone name: 315E02
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 197727 bases at least Q40
                Consensus quality: 198423 bases at least Q30
                Consensus quality: 198823 bases at least Q20
                Insert size: 202000; agarose-fp
                Insert size: 200249; sum-of-contigs
                Quality coverage: 9.87x in Q20 bases; agarose-fp
                Quality coverage: 9.95x in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2812: contig of 2812 bp in length
            *     2813     2912: gap of unknown length
            *     2913     9301: contig of 6389 bp in length
            *     9302     9401: gap of unknown length
            *     9402    17173: contig of 7772 bp in length
            *    17174    17273: gap of unknown length
            *    17274    26044: contig of 8771 bp in length
            *    26045    26144: gap of unknown length
            *    26145    65929: contig of 39785 bp in length
            *    65930    66029: gap of unknown length
            *    66030   109036: contig of 43007 bp in length
            *   109037   109136: gap of unknown length
            *   109137   200849: contig of 91713 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200849
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-315E2"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .2812
                     /note="assembly_fragment"
     misc_feature    2913. .9301
                     /note="assembly_fragment"
     misc_feature    9402. .17173
                     /note="assembly_fragment"
     misc_feature    17274. .26044
                     /note="assembly_fragment"
     misc_feature    26145. .65929
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    66030. .109036
                     /note="assembly_fragment"
     misc_feature    109137. .200849
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    53444 a  49164 c  49019 g  48614 t    608 others
ORIGIN


alignment_scores:
           Quality:   66.00            Length:     10
             Ratio:  7.333              Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-35 x AC091250   ..

 Align seg 1/1  to: AC091250  from: 1  to: 200849

       4 LysTrpProTrpTrpProTrpArgArgLys 13
         :::||||||||||||||||||||||   :::
  148453 AGGTGGCCTTGGTGGCCCTGGAGGTCTCGG 148482

seq_name: gb_htg:AC020365

seq_documentation_block:
LOCUS       AC020365    27780 bp    DNA             HTG       03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020365
VERSION     AC020365.1  GI:6664532
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 27780)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212917 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
```

```
          * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .27780
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     7869 a    6128 c    6190 g    7593 t
ORIGIN


alignment_scores:
            Quality:   65.00              Length:     15
              Ratio:  5.909                Gaps:      1
 Percent Similarity:  73.333   Percent Identity: 60.000

alignment_block:
 US-09-444-281-35 x AC020365   ..

 Align seg 1/1  to: AC020365  from: 1  to: 27780

       3 LysLysTrpPro...........TrpTrpProTrpArgArgLys 13
         |||:::||||||           ||||||||||||||:::|||
   10280 AAGAGGTGGCCCTGGAGGAAAAGGTGGTGGCCCTGGAGGAAGAAA 10324

seq_name: gb_in:AC008316

seq_documentation_block:
LOCUS        AC008316   160817 bp    DNA             INV       06-MAR-2001
DEFINITION   Drosophila melanogaster, chromosome 3R, region 85F-85F, BAC clone
             BACR23O04, complete sequence.
ACCESSION    AC008316
VERSION      AC008316.3  GI:13236627
KEYWORDS     HTG.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 160817)
  AUTHORS    Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
             Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
             Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
             Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
             Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
             Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
             Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
             Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
             McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
             Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
             Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
             Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
             Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE      Sequencing of Drosophila chromosome 3R, region 85F-85F
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 160817)
  AUTHORS    Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
             Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
             Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
             Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
             Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
             Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
             Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
             Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
             Rubin,G.M.
  TITLE      Direct Submission
  JOURNAL    Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
             Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT      On Mar 6, 2001 this sequence version replaced gi:6957904.
             Sequence submitted by:
             Berkeley Drosophila Genome Project
             Lawrence Berkeley National Laboratory, MS 64-121
             Berkeley, CA 94720
             This sequence was assembled using end sequences from a whole genome
             shotgun and from subclones of this BAC and its neighboring clones.
             For further information about this sequence, including its location
             and relationship to other sequences, please visit our sequence
             archive Web site (http://www.fruitfly.org/sequence/) or send email
             to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .160817
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="85F-85F"
                     /clone="BACR23O04 (D837)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT     45532 a   35217 c   34880 g   45188 t
ORIGIN


alignment_scores:
            Quality:   65.00              Length:     15
              Ratio:  5.909                Gaps:      1
 Percent Similarity:  73.333   Percent Identity: 60.000

alignment_block:
 US-09-444-281-35 x AC008316   ..

 Align seg 1/1  to: AC008316  from: 1  to: 160817

       3 LysLysTrpPro...........TrpTrpProTrpArgArgLys 13
         |||:::||||||           ||||||||||||||:::|||
   18897 AAGAGGTGGCCCTGGAGGAAAAGGTGGTGGCCCTGGAGGAAGAAA 18941

seq_name: gb_in:AC008315

seq_documentation_block:
LOCUS        AC008315   177028 bp    DNA             INV       17-FEB-2001
DEFINITION   Drosophila melanogaster, chromosome 3R, region 85E-85E, BAC clone
             BACR30F01, complete sequence.
ACCESSION    AC008315
VERSION      AC008315.9  GI:12957660
KEYWORDS     HTG.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 177028)
  AUTHORS    Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
             Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
             Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
             Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
             Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
             Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
             Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
             Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
             McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
             Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
             Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
             Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
             Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE      Sequencing of Drosophila chromosome 3R, region 85E-85E
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 177028)
  AUTHORS    Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
             Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
             Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
             Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
             Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
             Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
             Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
             Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
             Rubin,G.M.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 17, 2001 this sequence version replaced gi:7248925.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1. .177028
                      /organism="Drosophila melanogaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="3R"
                      /map="85E-85E"
                      /clone="BACR30F01 (D836)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanogaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    48567 a  39865 c  40667 g  47929 t
ORIGIN


alignment_scores:
            Quality:   65.00            Length:     15
              Ratio:  5.909              Gaps:      1
 Percent Similarity:  73.333   Percent Identity: 60.000

alignment_block:
 US-09-444-281-35 x AC008315   ..

 Align seg 1/1  to: AC008315  from: 1  to: 177028

      3 LysLysTrpPro.....TrpTrpProTrpArgArgLys 13
        |||:::||||||          ||||||||||||||:::|||
  126050 AAGAGGTGGCCCTGGAGGAAAAGGTGGTGGCCCTGGAGGAAGAAA 126094

seq_name: gb_in:AE003684

seq_documentation_block:
LOCUS       AE003684   219579 bp    DNA             INV       05-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386035 section 9
            of 105, complete sequence.
ACCESSION   AE003684 AE002708
VERSION     AE003684.2  GI:10726416
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 219579)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 219579)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7299232.
FEATURES              Location/Qualifiers
     source           1. .219579
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /chromosome="3R"
     mRNA             complement(join(608. .879,945. .1093,1160. .1371,1428. .1672,
                      1738. .2464,2521. .2728,2789. .3377,3441. .3601,4534. .5054))
                      /gene="CG9448"
                      /product="CT26774"
                      /db_xref="FLYBASE:FBan0009448"
                      /db_xref="FLYBASE:FBgn0037734"
     gene             complement(<608. .>5054)
                      /gene="CG9448"
                      /map="85E5-85E6"
                      /db_xref="FLYBASE:FBan0009448"
                      /db_xref="FLYBASE:FBgn0037734"
     CDS              complement(join(801. .879,945. .1093,1160. .1371,1428. .1672,
                      1738. .2464,2521. .2728,2789. .3377,3441. .3568))
                      /gene="CG9448"
                      /note="CG9448 gene product"
                      /codon_start=1
                      /db_xref="FLYBASE:FBan0009448"
                      /db_xref="FLYBASE:FBgn0037734"
                      /protein_id="AAF54429.1"
                      /db_xref="GI:7299233"
                      /translation="MCDTKDDAQKWKCETCTYENYPSSLKCTMCQASKPLLNEDIFRL
                      SPAQESCTVAEEAAAVEVAVMSPTPSSTCYSLQPQSQARQSNVADSEKWPCKVCTYLN
                      WPRSLRCVQCCTKRGGEAIERGKKDMDNEADGDRAGEALQALRISGSEENLANKPVQL
                      IGATASHRLSLSRGIDDATHLNNLANASHNQSQSQHRQPVLQQQMQLQLQPQQQRESS
                      SSAAVPPQQQKQCYVSKWACNSCTYENWPRSIKCSMCGKTREREISGSQNDLHASSSL
                      NSQEENQQQLQQPNVDTVSVNNSFNKKHIYQLGSSETINNCDTLQERQERRQRQIRRQ
                      VDWQWLNACLGVVENNYSAVEAYLSCGGNPARSLTSTEIAALNRNSAFDVGHTLIHLA
                      IRFHREEMLPMLLDQISGSGPGIKRVPSYVAPDLAADIRRHFANTLRLRKSGLPCHYV
                      QKHATFALPAEIEELPIPIQEQLYDELLDRDAQKQLETPPPALNWSLEITARLSSRMF
                      VLWNRSAGDCLLDSAMQATWGVFDRDNILRRALADTLHQCGHVFFTRWKEYEMLQASM
                      LHFTLEDSQFEEDWSTLLSLAGQPGSSLEQLHIFALAHILRRPIIVYGVKYVKSFRGE
                      DIGYARFEGVYLPLFWDQNFCTKSPIALGYTRGHFSALVPMEPFTRIDGRRDDVEDVT
                      YLPLMDCELKLLPIHFLTQSEVGNEESMMRQWLDVCVTDGGLLVAQQKLSKRPLLVAQ
                      MLEEWLNHYRRIAQVITAPFIRRPQITHYSSDGDSDEE"
     mRNA             join(5193. .5743,5803. .7634,7920. .8728)
                      /gene="dmt"
```

```
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT24655"
                     /db_xref="FLYBASE:FBan0008374"
                     /db_xref="FLYBASE:FBgn0016792"
gene                 <5193. .>8728
                     /gene="dmt"
                     /note="CG8374"
                     /map="85E5-85E6"
                     /db_xref="FLYBASE:FBan0008374"
                     /db_xref="FLYBASE:FBgn0016792"
CDS                  join(5263. .5743,5803. .7634,7920. .8180)
                     /gene="dmt"
                     /note="dmt gene product; Nucleotide sequence of the Celera
                     sequence differs from the published sequence for this
                     transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008374"
                     /db_xref="FLYBASE:FBgn0016792"
                     /protein_id="AAF54430.1"
                     /db_xref="GI:7299234"
                     /translation="MVRTRPVPCVPSPDVNTATRRNPGRPKKQSIGADLSTTISKPGR
                     PKKLSIGADLTTIRKPGRPKKLGADLTTIIRKPGRPTKLSNKQSLTALNEPECRIRKC
                     VVKIRRIKVKNGVVSEISRNDAPDVDASTFEPARNSTMYIPSKVQKSTQPQNIKENVA
                     GPETSPCQQRIRSKSPSGPAAEISPKKPPRFFHRDQPLTRTRSSVTRNVFDFLSQSQI
                     EDDSDREDPAADIIQRLVKDGKACVMVRSRKIGKPRAKRTAKKVRPVGNRRKVSTKDN
                     EPEPVKVVPKSIEPRLQKPSRSLSTIFEPEEDYSNDSEHGYVQPIEVPVQVHVEPSTS
                     KQAHEGAYSNLARSVMLNQTQAQNPLPSTDRRRELINMARQLVSTPLNRRAPPVTDVS
                     ATTTALSPIAHQSPNAVRTAGGKSPWRVSDDSPLPNTFMFGFNNSQMPSYSSDHVQRR
                     HVYVPDSPVEHAENIPHEESICPPLHEQNHDSNANDSNEENRPPPTISKSMSINDQES
                     EENAENFVHLPNPRRTLQKRTPFKDINILEVVTLPSWKKNVTATVSKEITPTRVAPRP
                     MATSSPAQRSQTRSNLFGFDDDFACEDIPRKSTTPSKGIAPTSMSLNREVTPALANRN
                     QTEVNTFGCNEILPCENIPKEKETNTEERTSSSQNLFGFDEFITESQDSPANFTGLSQ
                     NVNLHDKLHRLAELRPRDGELPQVSYTSIRSDCLGESHSKQTDIRYMLCSTMIAPPRP
                     AKRKPAAPTARESMGLFRVDQDEPEQSFADKQPRRTYVKERPQRKRKKRVQILYIESE
                     SEDEDEQDSHDKSLDSPEKKRHHVKRPRRDIEHEAKLEQFVTSFNKQCEEVEKFPVII
                     E"
mRNA                 complement(join(8675. .9217,9277. .9507,9574. .9721,
                     9777. .9983,10044. .10369,10427. .10503,10570. .12329,
                     12384. .12594,12684. .13156,13217. .13450,13530. .13674,
                     13773. .17273,17361. .17551,17614. .17852,17913. .18130,
                     18190. .18469,18533. .18730,19517. .19631))
                     /gene="hyd"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT26854"
                     /db_xref="FLYBASE:FBan0009484"
                     /db_xref="FLYBASE:FBgn0002431"
gene                 complement(<8675. .>19631)
                     /gene="hyd"
                     /note="CG9484"
                     /map="85E5-85E6"
                     /db_xref="FLYBASE:FBan0009484"
                     /db_xref="FLYBASE:FBgn0002431"
CDS                  complement(join(11575. .12329,12384. .12594,12684. .13156,
                     13217. .13450,13530. .13674,13773. .17273,17361. .17551,
                     17614. .17852,17913. .18130,18190. .18469,18533. .18730,
                     19517. .19569))
                     /gene="hyd"
                     /note="hyd gene product; Nucleotide sequence of the Celera
                     sequence differs from the published sequence for this
                     transcript".
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0009484"
                     /db_xref="FLYBASE:FBgn0002431"
                     /protein_id="AAF54431.1"
                     /db_xref="GI:7299235"
                     /translation="MQFVLQPLPGSDDQFIERIREVSDKVNRFGYGSHRIFEQLKIPV
                     KEVVIGPAHIGVLLEDGKAFRVSFSINSEKLDLTKSDAKCSTSGGSGTASASKAPSSS
                     RPMARSRARLLRATGRSNSTGQGSGSRSTGVIIGGSTSSRPLVTVPATYVPEELISQA
                     EVVLQGKSRNLIIRELQRTNLDVNLAVNNLLSRDDEEAEDTEEGADNYVPEDLISLLD
                     NGFSGDNNSVIIDPSDGLFSEEIFSNYSSIRNLLFDRIRSERSNANANAADSNQSTTR
                     STSSGTALTGNSGLSAQISVNADREAFSRWRDRQYYGPRRWISKDDYTWEKDADSKKK
                     EPSPMLSPIWISEELQPWPEKSSVRFKTIGALYSEFIALSESGDLYQWRWSDAEPYKS
                     ETENVYHPKTVSLNIVERVELISANFIRCSVVTETNRVATWMDEQLGYIGAKLEHSCC
                     AFNEFISDSITKIYVCSLYTVVKTESNNIYWWGVLPFDQRRFLWDKFRTKTKKPFKVV
                     ATDINVGAQVIMKKCPMYQSGSIGFTCSNGVPKVGQLLNSVWTFTDVCRMKIININTN
                     SGVDKSQAAGNNLNAHGITPDKDLPKSTAMPSTGSSKNGQSFSNSKESTDRIDMPPPP
                     SPASSTCSDTGSVTSHKRTKRATTKEDSNAPQEGRKDEELWEVKDVVFVEDKVGPVGK
                     VLKVDGDFVAVRFPAINAAAVAAAAAATSSTSNTASTSKEEGKEDDWQQCRLLRREDV
                     QIFRTAMSTRGPDWLQKQPKKINVGGDAAGAQILTLAVDSRGIHVIKKVLGKIHYSLY
                     NLYNCKQEQNCLFPTDCNSFIGSTPGNILMACNDDCSGNSSTIVLRDGNGALYPLAKD
                     CLGSIKDPQWFDLPPVKSITMSTISLPAMLSGVNLKSKVCMTALLFDTQKLMPHILRC
                     DVKNSFAALGRLEREDQADTALVVEERCDGARNIFHACVIMCAPSSNKDSPPDSPSGG
                     VEKKSLVGLSVARSIPTVSTSAYVSSIAFGASASSSNENSSFATMSSSAAGSASSTSR
                     DNRTNLRDMMNRLINSDQAEQSGSQPMATNNEDHAYIPWPAETPAASNLSASSSQNVS
                     DSIEDDISKIIPSSSQSSMLSNIKLGSPTYTFDLAQRREHALTILQQMCVSPALRPYL
                     CHMLSTKDAQGQTPFMLSVSCRAYEAGIILLNTILMLSEQDPQLKEAMIFPNGSPADQ
                     SPLHVICYNDTCSFTWTGADHINQNIFECKTCGLTGSLCCCTECARVCHKGHDCKLKR
                     TAPTAYCDCWEKCKCKALIAGNLTKRFALLCKLVSCTDLVTKFNSKGESILLFLIQTV
                     GRQIVEQRQYRFSVRVRNVSTAATGATGNNSVISNRKTSAAEIDNDMPDHDLEPPKFA
                     RKALERLLIDWNAVRSMIMSGAERGDVPNPAGSASENSNSEGFNMFIQTQHGSTLLDK
                     FTHSLIVKCTSDHLDTLLLTLVRELQNASVSNRSKEAEEVVRRFVRSVARVFVIFNLE
                     KQPNPEKRKSHSSCNKYVQSCVKVFQTLHKISIEELCEVSEALIAPVRLGVVRPTAPF
                     TMSSSNLDNSDDLFSVDPLAPSNVESPSEQILVHDAGNDQSANFNIQQNYDVVAMETI
                     RDASESEEVINREANSHNQDDELIENQRNEDGMQDDESDNDFTFNDAETESDSDDNQS
                     NQEVQRSVQAGATVGSENDIGVLFLEDESGDSSAQEEDGSEDGESDDQSDEFNFNEQQ
                     LERRSTNSNARSDLAPQTMQWAIRSRDTARSSVRVPTGSNMVFIDPMALRRSTVPAST
                     TVTTPSIEPHTMATTASNLARAFGITIRQISELISILSYNVLNDIETSLKIQNDEAIA
                     VQAFVEKRLKATWDWMFTVMDGTEAQLKFGAYLTNYTDPNHPLHPLNLSAQASSSQTP

alignment_scores:
            Quality:   65.00              Length:     15
              Ratio:  5.909                Gaps:      1
 Percent Similarity:  73.333   Percent Identity: 60.000

alignment_block:
 US-09-444-281-35 x AE003684   ..

 Align seg 1/1  to: AE003684  from: 1  to: 219579

       3 LysLysTrpPro.............TrpTrpProTrpArgArgLys 13
         |||:::|||||||             ||||||||||||||||:::||| 
   38362 AAGAGGTGGCCCTGGAGGAAAAGGTGGTGGCCCTGGAGGAAGAAA 38406

seq_name: gb_ro:MMA2IXCOA

seq_documentation_block:
LOCUS       MMA2IXCOA   19479 bp    DNA             ROD       08-APR-1994
DEFINITION  M.musculus alpha2 (IX) collagen gene, complete CDS.
ACCESSION   Z22923
VERSION     Z22923.1  GI:311949
KEYWORDS    alpha2 (IX) collagen.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 19479)
  AUTHORS   Perala,M., Elima,K., Metsaranta,M., Rosati,R., de Crombrugghe,B.
            and Vuorio,E.
  TITLE     The exon structure of the mouse alpha 2(IX) collagen gene shows
            unexpected divergence from the chick gene
  JOURNAL   J. Biol. Chem. 269 (7), 5064-5071 (1994)
  MEDLINE   94148964
REFERENCE   2  (bases 1 to 19479)
  AUTHORS   Perala,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1993) Merja Perala, Medical Biochemistry,
            University of Turku, Kiinamyllynkatu 10, 20520 Turku, Finland
FEATURES             Location/Qualifiers
     source          1. .19479
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="cMP9A2"
                     /clone_lib="Mouse genomic cosmid library (Clonetech)"
```

```
CAAT_signal     2600. .2604
exon            2677. .3131
                /gene="alpha2 (IX) collagen"
                /note="alternative transcription start site= 2558"
                /label=ex1
                /number=1
                /evidence=experimental
mRNA            join(2677. .3131,4311. .4385,5646. .5681,5768. .5830,
                7048. .7101,7214. .7249,7347. .7370,7630. .7683,8010. .8063,
                8168. .8215,8344. .8400,8485. .8538,8828. .8881,9268. .9321,
                9413. .9466,9572. .9625,12979. .13032,13393. .13446,
                13672. .13725,14530. .14574,14967. .15020,15284. .15337,
                15543. .15596,15674. .15745,15897. .15932,16069. .16113,
                16207. .16239,16375. .16521,16756. .16810,17086. .17274,
                17569. .17646,17912. .18617)
                /gene="alpha2 (IX) collagen"
gene            2677. .18617
                /gene="alpha2 (IX) collagen"
CDS             join(3060. .3131,4311. .4385,5646. .5681,5768. .5830,
                7048. .7101,7214. .7249,7347. .7370,7630. .7683,8010. .8063,
                8168. .8215,8344. .8400,8485. .8538,8828. .8881,9268. .9321,
                9413. .9466,9572. .9625,12979. .13032,13393. .13446,
                13672. .13725,14530. .14574,14967. .15020,15284. .15337,
                15543. .15596,15674. .15745,15897. .15932,16069. .16113,
                16207. .16239,16375. .16521,16756. .16810,17086. .17274,
                17569. .17646,17912. .18111)
                /gene="alpha2 (IX) collagen"
                /codon_start=1
                /product="alpha2 (IX) collagen"
                /protein_id="CAA80503.1"
                /db_xref="GI:311950"
                /db_xref="MGD:MGI:88466"
                /db_xref="SPTREMBL:Q07643"
                /translation="MTAVPAPRSLFVLLQVLWLALAQIRGPPGEPGPPGPPGPPGVPG
                SDGIDGDKGPPGKVGPPGSKGEPGKPGPDGPDGKPGIDGLMGAKGEPGPVGTPGVKGQ
                PGLPGPPGLPGPGFAGPPGPPGPVGLPGEIGTPGPKGDPGPEGPSGPPGPPGKPGRPG
                TIQGLEGSADFLCPTNCPAGVKGPQGLQGVKGHPGKRGILGDPGRQGKPGPKGDVGAS
                GEQGIPGPPGPQGIRGYPGMAGPKGEMGPRGYKGMVGSIGAAGPPGEEGPRGPPGEAG
                EKGDVGSQGARGPQGITGPKGITGPPGIDGKDGTPGIPGMKGSAGQVGRPGSPGHQGL
                AGVPGQPGTKGGPGDKGEPGQQGLPGVSGPPGKEGEPGPRGEIGPQGIMGQKGDQGER
                GPVGQPGPQGRQGPKGEQGPPGIPGPQGLPGIKGDKGSPGKTGPRGGVGDPGVAGLPG
                EKGEKGQSGEPGLKGQQGVRGETGYPGPSGDIGAPGVQGYPGLPGPRGLVGDRGVPGQ
                PGRQGVVGRAASDQHIVDVVLKMIQEQLAEVAVSAKREALGAAGMVGLPGPPGPPGYP
                GKQGPNGHPGPRGIPGIVGAVGQIGNTGPKGKRGEKGDRGEMGHGHPGMPGPPGIPGL
                PGRPGQAINGKDGDRGSPGAPGEAGRPGRPGPVGLPGFCEPAACLGASAYTSARLTEP
                GSIKGP"
exon            4311. .4385
                /gene="alpha2 (IX) collagen"
                /label=ex2
                /number=2
                /evidence=experimental
exon            5646. .5681
                /gene="alpha2 (IX) collagen"
                /label=ex3
                /number=3
                /evidence=experimental
exon            5768. .5830
                /gene="alpha2 (IX) collagen"
                /label=ex4
                /number=4
                /evidence=experimental
exon            7048. .7101
                /gene="alpha2 (IX) collagen"
                /label=ex5
                /number=5
                /evidence=experimental
exon            7214. .7249
                /gene="alpha2 (IX) collagen"
                /label=ex6
                /number=6
                /evidence=experimental
exon            7347. .7370
                /gene="alpha2 (IX) collagen"
                /label=ex7
                /number=7
                /evidence=experimental
exon            7630. .7683
                /gene="alpha2 (IX) collagen"
                /label=ex8
                /number=8
                /evidence=experimental
exon            8010. .8063
                /gene="alpha2 (IX) collagen"
                /label=ex9
                /number=9
                /evidence=experimental
exon            8168. .8215
                /gene="alpha2 (IX) collagen"
                /label=ex10
                /number=10
                /evidence=experimental
exon            8344. .8400
                /gene="alpha2 (IX) collagen"
                /label=ex11
                /number=11
                /evidence=experimental
exon            8485. .8538
                /gene="alpha2 (IX) collagen"
                /label=ex12
                /number=12
                /evidence=experimental
exon            8828. .8881
                /gene="alpha2 (IX) collagen"
                /label=ex13
                /number=13
                /evidence=experimental
exon            9268. .9321
                /gene="alpha2 (IX) collagen"
                /label=ex14
                /number=14
                /evidence=experimental
exon            9413. .9466
                /gene="alpha2 (IX) collagen"
                /label=ex15
                /number=15
                /evidence=experimental
exon            9572. .9625
                /gene="alpha2 (IX) collagen"
                /label=ex16
                /number=16
                /evidence=experimental
exon            12979. .13032
                /gene="alpha2 (IX) collagen"
                /label=ex17
                /number=17
                /evidence=experimental
exon            13393. .13446
                /gene="alpha2 (IX) collagen"
                /label=ex18
                /number=18
                /evidence=experimental
exon            13672. .13725
                /gene="alpha2 (IX) collagen"
                /label=ex19
                /number=19
                /evidence=experimental
exon            14530. .14574
                /gene="alpha2 (IX) collagen"
                /label=ex20
                /number=20
                /evidence=experimental
exon            14967. .15020
                /gene="alpha2 (IX) collagen"
                /label=ex21
                /number=21
                /evidence=experimental
```

```
     exon            15284. .15337
                     /gene="alpha2 (IX) collagen"
                     /label=ex22
                     /number=22
                     /evidence=experimental
     exon            15543. .15596
                     /gene="alpha2 (IX) collagen"
                     /label=ex23
                     /number=23
                     /evidence=experimental
     exon            15674. .15745
                     /gene="alpha2 (IX) collagen"
                     /label=ex24
                     /number=24
                     /evidence=experimental
     exon            15897. .15932
                     /gene="alpha2 (IX) collagen"
                     /label=ex25
                     /number=25
                     /evidence=experimental
     exon            16069. .16113
                     /gene="alpha2 (IX) collagen"
                     /label=ex26
                     /number=26
                     /evidence=experimental
     exon            16207. .16239

alignment_scores:
           Quality:   64.00          Length:      9
             Ratio:  8.000           Gaps:        0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-444-281-35 x MMA2IXCOA/rev

 Align seg 1/1  to reverse of: MMA2IXCOA  from: 1  to: 19479

       5 TrpProTrpTrpProTrpArgArgLys 13
         |||||||||||||||||||||   :::
    4355 TGGCCCTGGTGGCCCTGGAGGACCCGG 4329

seq_name: gb_pr:AL162423

seq_documentation_block:
LOCUS       AL162423    88323 bp    DNA             PRI       01-MAR-2001
DEFINITION  Human DNA sequence from clone RP11-456D21 on chromosome 9, complete
            sequence.
ACCESSION   AL162423
VERSION     AL162423.18  GI:13234851
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 88323)
  AUTHORS   Garner,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 5, 2001 this sequence version replaced gi:13184318.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-456D21 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-456D21 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-456D21 is at 1 in this sequence.
            The true left end of clone RP11-498E2 is at 88224 in this sequence.
            The true right end of clone RP11-297L11 is at 34844 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .88323
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-456D21"
                     /clone_lib="RPCI-11.2"
     repeat_region   362. .588
                     /note="HAL1 repeat: matches 595. .829 of consensus"
     repeat_region   693. .1009
                     /note="L1MEc repeat: matches 2085. .2405 of consensus"
     repeat_region   1168. .1751
                     /note="L1M4 repeat: matches 2630. .3230 of consensus"
     repeat_region   1775. .2047
                     /note="L2 repeat: matches 2248. .2526 of consensus"
     repeat_region   2061. .2215
                     /note="MIR repeat: matches 82. .252 of consensus"
     repeat_region   2443. .2788
                     /note="L1M4 repeat: matches 5399. .5754 of consensus"
     repeat_region   2889. .3184
                     /note="AluY repeat: matches 1. .298 of consensus"
     repeat_region   3306. .3446
                     /note="FLAM_C repeat: matches 2. .133 of consensus"
     repeat_region   3862. .4181
                     /note="AluJo repeat: matches 2. .312 of consensus"
     repeat_region   5285. .5366
                     /note="41 copies 2 mer aa 63% conserved"
     repeat_region   5681. .5983
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   6211. .6508
                     /note="AluJb repeat: matches 2. .298 of consensus"
     repeat_region   6566. .6961
                     /note="MLT1I repeat: matches 3. .402 of consensus"
     repeat_region   6989. .7284
                     /note="AluSg repeat: matches 1. .296 of consensus"
     repeat_region   8131. .8238
                     /note="MIR repeat: matches 124. .256 of consensus"
     repeat_region   8437. .8740
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   8936. .9019
                     /note="L2 repeat: matches 2660. .2749 of consensus"
     repeat_region   9150. .9379
                     /note="MIR repeat: matches 24. .259 of consensus"
     repeat_region   10513. .10672
                     /note="MIR repeat: matches 63. .218 of consensus"
     repeat_region   10674. .10795
                     /note="AluSg/x repeat: matches 172. .293 of consensus"
     repeat_region   11633. .11896
                     /note="L2 repeat: matches 2414. .2682 of consensus"
     repeat_region   11870. .12134
                     /note="MIR repeat: matches 8. .261 of consensus"
     repeat_region   12183. .12502
```

```
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   13079. .13193
                     /note="L2 repeat: matches 2613. .2746 of consensus"
     repeat_region   13236. .13416
                     /note="FRAM repeat: matches -4. .173 of consensus"
     repeat_region   13682. .13736
                     /note="MIR repeat: matches 93. .147 of consensus"
     repeat_region   13754. .14043
                     /note="L2 repeat: matches 2176. .2497 of consensus"
     repeat_region   14101. .14311
                     /note="L2 repeat: matches 2518. .2750 of consensus"
     repeat_region   15530. .15724
                     /note="MIR repeat: matches 30. .236 of consensus"
     repeat_region   15815. .15853
                     /note="L2 repeat: matches 2565. .2603 of consensus"
     misc_feature    16005. .16999
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   17381. .17689
                     /note="AluJb repeat: matches 6. .312 of consensus"
     repeat_region   17707. .18024
                     /note="AluSq repeat: matches 1. .313 of consensus"
     repeat_region   18047. .18554
                     /note="L1MB3 repeat: matches 5650. .6182 of consensus"
     repeat_region   18929. .19113
                     /note="MER5A repeat: matches 1. .188 of consensus"
     repeat_region   18962. .19120
                     /note="MER5A repeat: matches 9. .167 of consensus"
     repeat_region   19210. .19281
                     /note="L2 repeat: matches 1812. .1882 of consensus"
     repeat_region   19283. .19375
                     /note="AluY repeat: matches 217. .309 of consensus"
     repeat_region   19377. .19420
                     /note="22 copies 2 mer at 100% conserved"
     repeat_region   19421. .19563
                     /note="AluSg/x repeat: matches 123. .274 of consensus"
     repeat_region   19568. .19597
                     /note="15 copies 2 mer ga 93% conserved"
     repeat_region   20225. .20312
                     /note="L2 repeat: matches 2343. .2449 of consensus"
     repeat_region   20502. .20599
                     /note="L2 repeat: matches 2399. .2498 of consensus"
     repeat_region   21118. .21316
                     /note="L2 repeat: matches 2554. .2750 of consensus"
     repeat_region   21718. .21920
                     /note="MIR repeat: matches 33. .233 of consensus"
     repeat_region   22224. .22298
                     /note="L2 repeat: matches 2113. .2192 of consensus"
     repeat_region   22325. .22624
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   23006. .23138
                     /note="L2 repeat: matches 2340. .2474 of consensus"
     repeat_region   23362. .23636
                     /note="MIR repeat: matches 1. .261 of consensus"
     repeat_region   24001. .24137
                     /note="L1ME3A repeat: matches 5940. .6103 of consensus"
     repeat_region   24573. .24834
                     /note="AluSx repeat: matches 6. .268 of consensus"
     repeat_region   24835. .24874
                     /note="Alu repeat: matches 235. .275 of consensus"
     repeat_region   24906. .24963
                     /note="L2 repeat: matches 2129. .2187 of consensus"
     repeat_region   24995. .25190
                     /note="MIR repeat: matches 7. .220 of consensus"
     repeat_region   26116. .26287
                     /note="MER58A repeat: matches 43. .224 of consensus"
     repeat_region   26290. .26600
                     /note="AluSp repeat: matches 1. .312 of consensus"
     repeat_region   26605. .27039
                     /note="L1MB8 repeat: matches 5374. .5832 of consensus"
     repeat_region   27040. .27349
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   27351. .27811
                     /note="HAL1 repeat: matches 457. .940 of consensus"
     repeat_region   27839. .28045
                     /note="L1MB8 repeat: matches 5813. .6036 of consensus"
     repeat_region   28046. .28351
                     /note="AluSg repeat: matches 1. .296 of consensus"
     repeat_region   28352. .28494
                     /note="L1MB8 repeat: matches 6036. .6173 of consensus"
     repeat_region   28560. .28666
                     /note="MIR repeat: matches 34. .142 of consensus"
     repeat_region   28667. .28792
                     /note="FLAM_C repeat: matches 1. .125 of consensus"
     repeat_region   28793. .28919
                     /note="MIR repeat: matches 142. .262 of consensus"
     repeat_region   28887. .29074
                     /note="MIR repeat: matches 1. .179 of consensus"
     repeat_region   29183. .29224
                     /note="THE1C repeat: matches 328. .371 of consensus"
     repeat_region   29225. .29527
                     /note="AluSg repeat: matches 1. .303 of consensus"
     repeat_region   29528. .29874
                     /note="THE1C repeat: matches 1. .328 of consensus"
     repeat_region   30638. .31493
                     /note="L2 repeat: matches 1767. .2680 of consensus"
     repeat_region   33042. .33246
                     /note="L2 repeat: matches 2433. .2613 of consensus"

alignment_scores:
            Quality:   64.00              Length:     13
              Ratio:  5.818                Gaps:      0
 Percent Similarity:  84.615    Percent Identity: 53.846

alignment_block:
 US-09-444-281-35 x AL162423/rev    ..

 Align seg 1/1  to reverse of: AL162423  from: 1  to: 88323

       1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::::::::|||||||||||||   |||   ||||||
   47969 TTGGTGAGGAATTGGCCATGGTGGGCATGGAGCCGGAAG 47931

seq_name: gb_htg:AL356097_0

seq_documentation_block:
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AL356097 Accession AL356097
     Fragment Name      Begin     End
     AL356097_0             1     110000
     AL356097_1        100001     210000
     AL356097_2        200001     310000
     AL356097_3        300001     355026
LOCUS       AL356097   355026 bp    DNA              HTG       07-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-180A14, *** SEQUENCING IN
            PROGRESS ***, 38 unordered pieces.
ACCESSION   AL356097 AC058797
VERSION     AL356097.11  GI:11414557
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355026)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 15, 2001 this sequence version replaced gi:8077023
            gi:9797422.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
```

```
---------- Project Information
Center project name: bA180A14
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 43% of reads
Sequencing vector: plasmid; L08752; 56% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 333302 bases at least Q40
Consensus quality: 341682 bases at least Q30
Consensus quality: 346727 bases at least Q20
Insert size: 351326; sum-of-contigs
Insert size: 176896; 2.8% error; agarose-fp
Quality coverage: 3.95x in Q20 bases; sum-of-contigs Quality
coverage: 8.06x in Q20 bases; agarose-fp
----------
Draft Sequence Produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1    16446: contig of 16446 bp in length
*   16447 16546: gap of      100 bp
*   16547    22746: contig of 6200 bp in length
*   22747 22846: gap of      100 bp
*   22847    45482: contig of 22636 bp in length
*   45483 45582: gap of      100 bp
*   45583    49251: contig of 3669 bp in length
*   49252 49351: gap of      100 bp
*   49352    53239: contig of 3888 bp in length
*   53240 53339: gap of      100 bp
*   53340    58178: contig of 4839 bp in length
*   58179 58278: gap of      100 bp
*   58279    71754: contig of 13476 bp in length
*   71755 71854: gap of      100 bp
*   71855    76641: contig of 4787 bp in length
*   76642 76741: gap of      100 bp
*   76742    80415: contig of 3674 bp in length
*   80416 80515: gap of      100 bp
*   80516    89902: contig of 9387 bp in length
*   89903 90002: gap of      100 bp
*   90003    93661: contig of 3659 bp in length
*   93662 93761: gap of      100 bp
*   93762   101785: contig of 8024 bp in length
*  101786 101885: gap of      100 bp
*  101886   109333: contig of 7448 bp in length
*  109334 109433: gap of      100 bp
*  109434   121288: contig of 11855 bp in length
*  121289 121388: gap of      100 bp
*  121389   128767: contig of 7379 bp in length
*  128768 128867: gap of      100 bp
*  128868   131917: contig of 3050 bp in length
*  131918 132017: gap of      100 bp
*  132018   138795: contig of 6778 bp in length
*  138796 138895: gap of      100 bp
*  138896   144660: contig of 5765 bp in length
*  144661 144760: gap of      100 bp
*  144761   156587: contig of 11827 bp in length
*  156588 156687: gap of      100 bp
*  156688   158939: contig of 2252 bp in length
*  158940 159039: gap of      100 bp
*  159040   161353: contig of 2314 bp in length
*  161354 161453: gap of      100 bp
*  161454   165170: contig of 3717 bp in length
*  165171 165270: gap of      100 bp
*  165271   168479: contig of 3209 bp in length
*  168480 168579: gap of      100 bp
*  168580   172188: contig of 3609 bp in length
*  172189 172288: gap of      100 bp
*  172289   177084: contig of 4796 bp in length
*  177085 177184: gap of      100 bp
*  177185   199876: contig of 22692 bp in length
*  199877 199976: gap of      100 bp
*  199977   222186: contig of 22210 bp in length
*  222187 222286: gap of      100 bp
*  222287   240235: contig of 17949 bp in length
*  240236 240335: gap of      100 bp
*  240336   263715: contig of 23380 bp in length
*  263716 263815: gap of      100 bp
*  263816   269270: contig of 5455 bp in length
*  269271 269370: gap of      100 bp
*  269371   271745: contig of 2375 bp in length
*  271746 271845: gap of      100 bp
*  271846   277024: contig of 5179 bp in length
*  277025 277124: gap of      100 bp
*  277125   280037: contig of 2913 bp in length
*  280038 280137: gap of      100 bp
*  280138   307328: contig of 27191 bp in length
*  307329 307428: gap of      100 bp
*  307429   313985: contig of 6557 bp in length
*  313986 314085: gap of      100 bp
*  314086   330085: contig of 16000 bp in length
*  330086 330185: gap of      100 bp
*  330186   334887: contig of 4702 bp in length
*  334888 334987: gap of      100 bp
*  334988   355026: contig of 20039 bp in length.
FEATURES             Location/Qualifiers
     source          1. .355026
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-180A14"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .16446
                     /note="assembly_fragment:02644
                     fragment_chain:1"
     misc_feature    16547. .22746
                     /note="assembly_fragment:02589
                     fragment_chain:1"
     misc_feature    22847. .45482
                     /note="assembly_fragment:03035
                     fragment_chain:1"
     misc_feature    45583. .49251
                     /note="assembly_fragment:00032"
     misc_feature    49352. .53239
                     /note="assembly_fragment:00060"
     misc_feature    53340. .58178
                     /note="assembly_fragment:00105"
     misc_feature    58279. .71754
                     /note="assembly_fragment:00250"
     misc_feature    71855. .76641
                     /note="assembly_fragment:00381"
     misc_feature    76742. .80415
                     /note="assembly_fragment:00389"
     misc_feature    80516. .89902
                     /note="assembly_fragment:00591"
     misc_feature    90003. .93661
                     /note="assembly_fragment:00750"
     misc_feature    93762. .101785
                     /note="assembly_fragment:00754"
     misc_feature    101886. .109333
                     /note="assembly_fragment:00819"
     misc_feature    109434. .121288
                     /note="assembly_fragment:00889"
     misc_feature    121389. .128767
                     /note="assembly_fragment:00932"
     misc_feature    128868. .131917
                     /note="assembly_fragment:00972"
     misc_feature    132018. .138795
                     /note="assembly_fragment:01033"
```

```
     misc_feature    138896. .144660
                     /note="assembly_fragment:01363"
     misc_feature    144761. .156587
                     /note="assembly_fragment:01542"
     misc_feature    156688. .158939
                     /note="assembly_fragment:01630"
     misc_feature    159040. .161353
                     /note="assembly_fragment:01684"
     misc_feature    161454. .165170
                     /note="assembly_fragment:01912"
     misc_feature    165271. .168479
                     /note="assembly_fragment:01914"
     misc_feature    168580. .172188
                     /note="assembly_fragment:02231"
     misc_feature    172289. .177084
                     /note="assembly_fragment:02455"
     misc_feature    177185. .199876
                     /note="assembly_fragment:02553"
     misc_feature    199977. .222186
                     /note="assembly_fragment:02605"
     misc_feature    222287. .240235
                     /note="assembly_fragment:02624"
     misc_feature    240336. .263715
                     /note="assembly_fragment:03030"

alignment_scores:
           Quality:   64.00          Length:      9
             Ratio:  7.111            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-35 x AL356097_0/rev   ..

 Align seg 1/1  to reverse of: AL356097_0  from: 1  to: 110000

       5 TrpProTrpTrpProTrpArgArgLys 13
         ||||||||||||||||||||::::::
    8764 TGGCCCTGGTGGCCTTGGAGGGAGGAG 8738

seq_name: gb_htg:AC027113

seq_documentation_block:
LOCUS       AC027113   159391 bp    DNA             HTG       25-MAY-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-672I2 map 5, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC027113
VERSION     AC027113.3  GI:8081286
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159391)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 5, clone RP11-672I2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159391)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7656748.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8843
                Center clone name: 672_I_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149154 bases at least Q40
                Consensus quality: 154165 bases at least Q30
                Consensus quality: 156049 bases at least Q20
                Insert size: 195000; agarose-fp
                Insert size: 157491; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1373: contig of 1373 bp in length
            *     1374 1473: gap of      100 bp
            *     1474     3371: contig of 1898 bp in length
            *     3372 3471: gap of      100 bp
            *     3472     5991: contig of 2520 bp in length
            *     5992 6091: gap of      100 bp
            *     6092     8001: contig of 1910 bp in length
            *     8002 8101: gap of      100 bp
            *     8102    10452: contig of 2351 bp in length
            *    10453 10552: gap of      100 bp
            *    10553    13499: contig of 2947 bp in length
            *    13500 13599: gap of      100 bp
            *    13600    17300: contig of 3701 bp in length
            *    17301 17400: gap of      100 bp
            *    17401    23167: contig of 5767 bp in length
            *    23168 23267: gap of      100 bp
            *    23268    27394: contig of 4127 bp in length
            *    27395 27494: gap of      100 bp
            *    27495    31941: contig of 4447 bp in length
            *    31942 32041: gap of      100 bp
            *    32042    36867: contig of 4826 bp in length
            *    36868 36967: gap of      100 bp
            *    36968    43237: contig of 6270 bp in length
            *    43238 43337: gap of      100 bp
            *    43338    52795: contig of 9458 bp in length
            *    52796 52895: gap of      100 bp
            *    52896    61446: contig of 8551 bp in length
            *    61447 61546: gap of      100 bp
            *    61547    68822: contig of 7276 bp in length
            *    68823 68922: gap of      100 bp
            *    68923    79907: contig of 10985 bp in length
            *    79908 80007: gap of      100 bp
            *    80008    91003: contig of 10996 bp in length
```

```
            *    91004 91103: gap of      100 bp
            *    91104   109428: contig of 18325 bp in length
            *   109429 109528: gap of      100 bp
            *   109529   135277: contig of 25749 bp in length
            *   135278 135377: gap of      100 bp
            *   135378   159391: contig of 24014 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159391
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5"
                     /clone="RP11-672I2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1373
                     /note="assembly_fragment"
     misc_feature    1474. .3371
                     /note="assembly_fragment"
     misc_feature    3472. .5991
                     /note="assembly_fragment"
     misc_feature    6092. .8001
                     /note="assembly_fragment"
     misc_feature    8102. .10452
                     /note="assembly_fragment"
     misc_feature    10553. .13499
                     /note="assembly_fragment"
     misc_feature    13600. .17300
                     /note="assembly_fragment"
     misc_feature    17401. .23167
                     /note="assembly_fragment"
     misc_feature    23268. .27394
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    27495. .31941
                     /note="assembly_fragment"
     misc_feature    32042. .36867
                     /note="assembly_fragment"
     misc_feature    36968. .43237
                     /note="assembly_fragment"
     misc_feature    43338. .52795
                     /note="assembly_fragment"
     misc_feature    52896. .61446
                     /note="assembly_fragment"
     misc_feature    61547. .68822
                     /note="assembly_fragment"
     misc_feature    68923. .79907
                     /note="assembly_fragment"
     misc_feature    80008. .91003
                     /note="assembly_fragment"
     misc_feature    91104. .109428
                     /note="assembly_fragment"
     misc_feature    109529. .135277
                     /note="assembly_fragment"
     misc_feature    135378. .159391
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    43727 a  35737 c  36282 g  41744 t   1901 others
ORIGIN


alignment_scores:
           Quality:   64.00          Length:      9
             Ratio:  7.111            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-35 x AC027113   ..

 Align seg 1/1  to: AC027113  from: 1  to: 159391

       5 TrpProTrpTrpProTrpArgArgLys 13
```

```
                 ||||||||||||||||||||:::::
   87682 TGGCCCTGGTGGCCTTGGAGGGAGGAG 87708

seq_name: gb_htg:AL358473

seq_documentation_block:
LOCUS       AL358473   187845 bp    DNA              HTG       07-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-168O16, *** SEQUENCING IN
            PROGRESS ***, 7 unordered pieces.
ACCESSION   AL358473 AC027712
VERSION     AL358473.8  GI:11322009
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187845)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 15, 2001 this sequence version replaced gi:7712176
            gi:10800605.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA168O16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 36% of reads
            Sequencing vector: plasmid; L08752; 63% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 186132 bases at least Q40
            Consensus quality: 186498 bases at least Q30
            Consensus quality: 186851 bases at least Q20
            Insert size: 187245; sum-of-contigs
            Insert size: 186146; agarose-fp
            Quality coverage: 10.30x in Q20 bases; sum-of-contigs Quality
            coverage: 10.58x in Q20 bases; agarose-fp
            --------------
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    11961: contig of 11961 bp in length
            *   11962 12061: gap of      100 bp
            *   12062    95058: contig of 82997 bp in length
            *   95059 95158: gap of      100 bp
            *   95159   124623: contig of 29465 bp in length
            *  124624 124723: gap of      100 bp
            *  124724   129609: contig of 4886 bp in length
            *  129610 129709: gap of      100 bp
            *  129710   151381: contig of 21672 bp in length
            *  151382 151481: gap of      100 bp
            *  151482   174927: contig of 23446 bp in length
            *  174928 175027: gap of      100 bp
            *  175028   187845: contig of 12818 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187845
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /chromosome="1"
                     /clone="RP11-168O16"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .11961
                     /note="assembly_fragment:01682
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    12062. .95058
                     /note="assembly_fragment:01450
                     fragment_chain:1"
     misc_feature    95159. .124623
                     /note="assembly_fragment:01175
                     fragment_chain:1"
     misc_feature    124724. .129609
                     /note="assembly_fragment:04318
                     fragment_chain:1"
     misc_feature    129710. .151381
                     /note="assembly_fragment:00767
                     fragment_chain:2"
     misc_feature    151482. .174927
                     /note="assembly_fragment:01178
                     fragment_chain:2"
     misc_feature    175028. .187845
                     /note="assembly_fragment:02771
                     clone_end:T7
                     vector_side:right"
BASE COUNT    46168 a  48356 c  47958 g  44763 t    600 others
ORIGIN


alignment_scores:
            Quality:   64.00              Length:      9
              Ratio:  7.111               Gaps:        0
 Percent Similarity: 100.000  Percent Identity: 77.778

alignment_block:
 US-09-444-281-35 x AL358473    ..

 Align seg 1/1  to: AL358473  from: 1  to: 187845

       5 TrpProTrpTrpProTrpArgArgLys 13
         |||||||||||||||||||||::::::
  133213 TGGCCCTGGTGGCCTTGGAGGGAGGAG 133239
```

```
OM of: US-09-444-281-35 to: N_Geneseq_1101:*   out_format : pfs

Date: Jan 4, 2002 11:02 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084144_16249/app_query.fasta_1.210
 -DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09444281_@CGN1_1_267 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-35
Query length: 13
Database: N_Geneseq_1101:*
Database sequences: 930621
Database length: 428662619
Search time (sec): 286.760000



score_list:
Sequence                                  Strd Orig    %Score   EScore Len  ! Documentation  ..
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV60908 +    91.00  213.55  0.0008    88  ! D
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27291 +    91.00  211.37  0.0011   114  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27296 +    76.00  179.37  0.0648   108  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27298 +    76.00  178.92  0.0687   114  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27294 +    76.00  176.55  0.0931   151  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV83788 +    73.00  181.45  0.0497    39  ! A
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29389 +    73.00  179.88  0.0608    47  ! E
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29390 -    73.00  179.88  0.0608    47  ! E
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29364 +    73.00  178.87  0.0692    53  ! A
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ40246 +    73.00  176.65  0.0919    69  ! C
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ49764 +    73.00  167.25  0.3069   211  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ45123 +    73.00  167.25  0.3069   211  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ20646 +    73.00  138.50  12.26    6446  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF82334 +    73.00  138.50  12.26    6446  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28519 +    66.00  152.27  2.10      207  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28510 +    64.00  161.98  0.6036    39  ! R
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27290 -    62.00  155.90  1.32      48  ! A
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27287 +    62.00  155.73  1.34      49  ! G
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28512 -    62.00  155.23  1.43      52  ! R
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28514 -    62.00  155.23  1.43      52  ! R
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF60750 +    62.00  130.13  35.87     1029  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT84485 -    60.00  114.32  272.47    4031  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ46155 -    59.00  119.53  139.56    1676  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ46160 -    59.00  117.60  178.91    2110  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT06037 +    58.00  118.25  164.48    1509  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT87878 +    58.00  118.25  164.48    1509  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT14643 +    58.00  117.78  174.73    1596  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ66634 +    58.00  115.96  220.83    1983  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ09841 +    58.00  115.95  220.95    1984  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH65605 +    58.00  111.02  415.97    3567  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH68525 +    58.00   72.45  5.8e+04  349980
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI27362 +    57.50  127.95  47.45     419  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI56255 +    57.50  127.95  47.45     419  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI18230 +    57.50  125.01  69.13     594  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI43269 +    57.50  125.01  69.13     594  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT64960 +    57.50  115.26  241.44    1894  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27745 +    57.50  115.26  241.44    1894  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ83735 +    57.50  104.36  976.92    6922  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ84662 +    57.50  104.23  993.67    7032  !
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA71720 +    57.50  104.23  993.67    7032  !
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ84663 +    57.50  104.16  1.0e+03   7089
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA71721 +    57.50  104.16  1.0e+03   7089
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC85513 +    57.00  109.18  526.46    3431
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC85514 +    57.00  109.09  532.59    3468
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX06943 +    56.00  120.59  121.85     683

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV60908

seq_documentation_block:
ID   AAV60908 standard; DNA; 88 BP.
XX
AC   AAV60908;
XX
DT   11-JAN-1999  (first entry)
XX
DE   DNA fragment encoding MBI11.
XX
KW   MBI28; cationic peptide; plasmid pKL1; small cryptic plasmid;
KW   replication; RepA; vector; RAMP; human; MBI11; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9841636-A2.
XX
PD   24-SEP-1998.
XX
PF   16-MAR-1998;   98WO-CA00214.
XX
PR   14-MAR-1997;   97US-0040722.
XX
PA   (BURI/) BURIAN J.
PA   (KAYW/) KAY W W.
XX
PI   Burian J,  Kay WW;
XX
DR   WPI; 1998-531571/45.
XX
PT   Increasing plasmid copy number in a cell with the repA gene product
PT   - and an small cryptic plasmid ori sequence, useful for high level
PT   expression of e.g. cytokines, antigens or therapeutic proteins
XX
PS   Example 16; Page 57; 82pp; English.
XX
CC   This oligonucleotide was used as a template in a PCR reaction (see
CC   also AAV60909-10) to generate a DNA fragment encoding the cationic
CC   peptide MBI11 (see AAW71690).  The PCR product was cloned into
CC   the universal vector pR2h-B1, which contains the R21 replication
CC   leader of RepA (see AAW71686) and 2 tandem copies of the prepro
CC   region (Hpro) of human defensin.  The vector provides expression
CC   of R21-hpro-MBI11 fusion in host cells.  The invention provides
CC   controlled replication plasmid vectors (RAMP vectors) comprising a
CC   replication origin of a small cryptic plasmid such as pKL1 (see
CC   AAV58292) and a gene encoding RepA (see AAW71686).  The vectors can
CC   reach very high levels of plasmid replication, but are not lethal
CC   to the host cell, and can be used to direct the high level
CC   expression of e.g. cytokines, antigens and therapeutic proteins.
XX
SQ   Sequence 88 BP; 20 A; 18 C; 25 G; 25 T; 0 other;


alignment_scores:
           Quality:   91.00              Length:     13
             Ratio:  7.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAV60908   ..

 Align seg 1/1  to: AAV60908  from: 1  to: 88

       1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         |||||||||||||||||||||||||||||||||||||||
```

```
      25 ATCCTGAAAAAATGGCCGTGGTGGCCGTGGCGTCGTAAA 63

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27291

seq_documentation_block:
ID   AAA27291 standard; DNA; 114 BP.
XX
AC   AAA27291;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI-11 fragment.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 4; Page 37; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11 fragment. MBI-11 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 114 BP; 25 A; 26 C; 30 G; 33 T; 0 other;


alignment_scores:
            Quality:   91.00             Length:     13
              Ratio:  7.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAA27291

 Align seg 1/1  to: AAA27291  from: 1  to: 114

       1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         |||||||||||||||||||||||||||||||||||||||
      41 ATCCTGAAAAAATGGCCGTGGTGGCCGTGGCGTCGTAAA 79

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27296

seq_documentation_block:
ID   AAA27296 standard; DNA; 108 BP.
XX
AC   AAA27296;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI 2X11B7 poly cassette.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 5; Page 39; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11B7 fragment. This fragment was used in the expression
CC   cassette. MBI-11B7 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 108 BP; 18 A; 33 C; 31 G; 26 T; 0 other;


alignment_scores:
            Quality:   76.00             Length:     13
              Ratio:  6.333               Gaps:      0
 Percent Similarity:  92.308   Percent Identity: 69.231

alignment_block:
 US-09-444-281-35 x AAA27296

 Align seg 1/1  to: AAA27296  from: 1  to: 108

       1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::   :::||||||||||||||||||||||||||
      38 ATGATTCTGCGTTGGCCGTGGTGGCCGTGGCGTCGCAAA 76

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27298

seq_documentation_block:
ID   AAA27298 standard; DNA; 114 BP.
XX
AC   AAA27298;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI 11B7 first cassette.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
```

```
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 5; Page 40; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11B7 fragment. This fragment was used in the expression
CC   cassette. MBI-11B7 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 114 BP; 20 A; 34 C; 32 G; 28 T; 0 other;


alignment_scores:
            Quality:   76.00              Length:     13
              Ratio:  6.333                Gaps:      0
 Percent Similarity:  92.308   Percent Identity: 69.231

alignment_block:
 US-09-444-281-35 x AAA27298   ..

 Align seg 1/1  to: AAA27298  from: 1  to: 114

       1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::   :::||||||||||||||||||||||||||||
      44 ATGATTCTGCGTTGGCCGTGGTGGCCGTGGCGTCGCAAA 82

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27294

seq_documentation_block:
ID   AAA27294 standard; DNA; 51 BP.
XX
AC   AAA27294;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Oligonucleotide used for synthesis of MBI 2X11B7 last cassette.
XX
KW   Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW   MBI-11; indolicidin; bovine; ss.
XX
OS   Synthetic.
XX
PN   WO200031279-A2.
XX
PD   02-JUN-2000.
XX
PF   19-NOV-1999;   99WO-CA01107.
XX
PR   20-NOV-1998;   98US-0109218.
XX
PA   (MICR-) MICROLOGIX BIOTECH INC.
XX
PI   Burian J,  Bartfeld D;
XX
DR   WPI; 2000-400086/34.
XX
PT   Multi-domain fusion protein expression cassette used for high yield
PT   stable production of foreign peptide gene products -
XX
PS   Example 5; Page 38; 73pp; English.
XX
CC   A novel method allows the efficient production of cationic peptides in
CC   recombinant host cells. The method involves construction of a
CC   multi-domain fusion protein expression cassette comprising a promoter and
CC   a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC   of anionic peptide sequences in the linker sequences neutralises the
CC   positive charge of the cationic peptide so that the charge of the
CC   fusion protein is controlled. This cassette allows high yield, stable
CC   production of the cationic peptide. Cationic peptides such as
CC   bovine indolicidin may be used as antimicrobial agents. The present
CC   sequence is an oligonucleotide that was used to synthesise a
CC   MBI-11B7 fragment. This fragment was used in the expression
CC   cassette. MBI-11B7 is a cationic peptide derived from modifications
CC   of indolicidin.
XX
SQ   Sequence 151 BP; 22 A; 44 C; 49 G; 36 T; 0 other;


alignment_scores:
            Quality:   76.00              Length:     13
              Ratio:  6.333                Gaps:      0
 Percent Similarity:  92.308   Percent Identity: 69.231

alignment_block:
 US-09-444-281-35 x AAA27294   ..

 Align seg 1/1  to: AAA27294  from: 1  to: 151

       1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::   :::||||||||||||||||||||||||||||
      38 ATGATTCTGCGTTGGCCGTGGTGGCCGTGGCGTCGCAAA 76

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV83788

seq_documentation_block:
ID   AAV83788 standard; DNA; 39 BP.
XX
AC   AAV83788;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Antimicrobial peptide Indolicidin encoding DNA.
XX
KW   Antimicrobial; fusion; acidic peptide; recombinant; microorganism;
KW   guamerin; basic peptide; Indolicidin; ss.
XX
OS   Synthetic.
OS   Bos sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..39
FT                   /*tag=  a
FT                   /note= "the start and stop codons are not indicated"
XX
PN   WO9854336-A1.
XX
PD   03-DEC-1998.
XX
PF   28-MAY-1998;   98WO-KR00132.
```

```
XX
PR   09-APR-1998;   98KR-0013372.
PR   28-MAY-1997;   97KR-0021312.
XX
PA   (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
PA   (SAMY-) SAMYANG GENEX CORP.
XX
PI   Hong S,  Kang MH,  Kim JH,  Kim S,  Lee H,  Lee JH;
XX
DR   WPI; 1999-059844/05.
DR   P-PSDB; AAW87609.
XX
PT   New method for mass production of antimicrobial peptides - by
PT   constructing fusion genes comprising acidic and antimicrobial
PT   peptide genes and transforming host with vector containing these
XX
PS   Example 6; Page 18; 52pp; English.
XX
CC   The invention relates to mass production of antimicrobial peptides. The
CC   method comprises constructing a fusion gene containing a first gene
CC   encoding a negatively charged acidic peptide having at least two cysteine
CC   residues, and a second gene encoding a positively charged basic
CC   antimicrobial peptide. A host microorganism is transformed with a vector
CC   containing the fusion gene and then cultured. The expressed antimicrobial
CC   peptide is then recovered. The method is used to mass produce
CC   antimicrobial peptides in recombinant microorganisms. The inhibitory
CC   effect of the expressed antimicrobial peptide upon the growth of the host
CC   microorganism is considerably reduced by fusing it to the acidic peptide.
CC   Therefore, the use of the fusion gene provides an economic, recombinant
CC   alternative of mass producing antimicrobial peptides, which overcomes the
CC   disadvantages of low-productivity and poor economy, previously
CC   encountered by recombinant and chemical methods. The present sequence
CC   represents the DNA encoding an antimicrobial peptide Indolicidin. This
CC   can be used along with the acidic peptide Guamerin gene in the
CC   construction of the fusion gene.
XX
SQ   Sequence 39 BP; 4 A; 10 C; 16 G; 9 T; 0 other;


alignment_scores:
           Quality:   73.00            Length:      9
             Ratio:  8.111              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAV83788   ..

 Align seg 1/1  to: AAV83788  from: 1  to: 39

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||||
      13 AAATGGCCGTGGTGGCCGTGGCGTCGT 39

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29389

seq_documentation_block:
ID   AAZ29389 standard; DNA; 47 BP.
XX
AC   AAZ29389;
XX
DT   29-FEB-2000  (first entry)
XX
DE   PCR primer-15 for synthesis of antimicrobial peptide Indolicidin gene.
XX
KW   PCR primer; anti-microbial peptide; Indolicidin gene; DNA construct;
KW   glutamine pyrophosphoribosyl pyrophosphate amidotransferase gene;
KW   purF gene; fusion peptide; mass production; pharmaceutical industry;
KW   food industry; ss.
XX
OS   Synthetic.
XX
PN   WO9964611-A1.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999;   99WO-KR00282.
XX
PR   09-JUN-1998;   98KR-0022117.
PR   14-MAY-1999;   99KR-0017920.
XX
PA   (SAMY-) SAMYANG GENEX CORP.
XX
PI   Kim JH,  Kang MH,  Lee J,  Park SH,  Lee JW,  Hong SS,  Lee H;
XX
DR   WPI; 2000-097542/08.
XX
PT   New DNA constructs useful for mass production of antimicrobial peptides
PT   in microorganism hosts  -
XX
PS   Example 1; Page 13; 67pp; English.
XX
CC   The present sequence is a chemically synthesised PCR primer which was
CC   used to synthesise a gene encoding antimicrobial peptide Indolicidin.
CC   The antimicrobial peptide gene is used in a DNA construct that comprises
CC   entire, partial or a derivative of purF gene (glutamine
CC   pyrophosphoribosyl pyrophosphate amidotransferase gene). The DNA
CC   construct allows mass production of the antimicrobial peptide in
CC   microbial hosts without killing the host cells. The antimicrobial
CC   peptides are useful commercially in the pharmaceutical and
CC   food industries.
XX
SQ   Sequence 47 BP; 6 A; 11 C; 19 G; 11 T; 0 other;


alignment_scores:
           Quality:   73.00            Length:      9
             Ratio:  8.111              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAZ29389   ..

 Align seg 1/1  to: AAZ29389  from: 1  to: 47

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||||
      17 AAATGGCCGTGGTGGCCGTGGCGTCGT 43

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29390

seq_documentation_block:
ID   AAZ29390 standard; DNA; 47 BP.
XX
AC   AAZ29390;
XX
DT   29-FEB-2000  (first entry)
XX
DE   PCR primer-16 for synthesis of antimicrobial peptide Indolicidin gene.
XX
KW   PCR primer; anti-microbial peptide; Indolicidin gene; DNA construct;
KW   glutamine pyrophosphoribosyl pyrophosphate amidotransferase gene;
KW   purF gene; fusion peptide; mass production; pharmaceutical industry;
KW   food industry; ss.
XX
OS   Synthetic.
XX
PN   WO9964611-A1.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999;   99WO-KR00282.
XX
PR   09-JUN-1998;   98KR-0022117.
PR   14-MAY-1999;   99KR-0017920.
XX
PA   (SAMY-) SAMYANG GENEX CORP.
```

```
XX
PI   Kim JH,  Kang MH,  Lee J,  Park SH,  Lee JW,  Hong SS,  Lee H;
XX
DR   WPI; 2000-097542/08.
XX
PT   New DNA constructs useful for mass production of antimicrobial peptides
PT   in microorganism hosts  -
XX
PS   Example 1; Page 13; 67pp; English.
XX
CC   The present sequence is a chemically synthesised PCR primer which was
CC   used to synthesise a gene encoding antimicrobial peptide Indolicidin.
CC   The antimicrobial peptide gene is used in a DNA construct that comprises
CC   entire, partial or a derivative of purF gene (glutamine
CC   pyrophosphoribosyl pyrophosphate amidotransferase gene). The DNA
CC   construct allows mass production of the antimicrobial peptide in
CC   microbial hosts without killing the host cells. The antimicrobial
CC   peptides are useful commercially in the pharmaceutical and
CC   food industries.
XX
SQ   Sequence 47 BP; 12 A; 18 C; 10 G; 7 T; 0 other;


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAZ29390/rev  ..

 Align seg 1/1  to reverse of: AAZ29390  from: 1  to: 47

       4 LysTrpProTrpTrpProTrpArgArg 12
         |||||||||||||||||||||||||||
      35 AAATGGCCGTGGTGGCCGTGGCGTCGT 9

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ29364

seq_documentation_block:
ID   AAZ29364 standard; DNA; 53 BP.
XX
AC   AAZ29364;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Antimicrobial peptide, Indolicidin encoding DNA.
XX
KW   purF gene; glutamine pyrophosphoribosyl pyrophosphate amidotransferase;
KW   purF derivative; fusion partner; antimicrobial peptide; Indolicidin;
KW   mass production; cleavage site; hydroxylamine; CNBr; DNA construct; cow;
KW   neutralise; toxicity; pharmaceutical industry; food industry; ds.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             5..46
FT                   /*tag= a
FT                   /product= "Indolicidin peptide"
FT                   /note= "Antimicrobial peptide used in DNA construct"
XX
PN   WO9964611-A1.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999;   99WO-KR00282.
XX
PR   09-JUN-1998;   98KR-0022117.
PR   14-MAY-1999;   99KR-0017920.
XX
PA   (SAMY-) SAMYANG GENEX CORP.
XX
PI   Kim JH,  Kang MH,  Lee J,  Park SH,  Lee JW,  Hong SS,  Lee H;
XX
DR   WPI; 2000-097542/08.
DR   P-PSDB; AAY44324.
XX
PT   New DNA constructs useful for mass production of antimicrobial peptides
PT   in microorganism hosts  -
XX
PS   Claim 1; Fig 1; 67pp; English.
XX
CC   The present DNA sequence encodes an antimicrobial peptide, Indolicidin
CC   derived from cow, Bos taurus. It is used along with a derivative
CC   of purF gene sequence that functions as a fusion partner.
CC   A DNA construct that comprises, this antimicrobial peptide encoding
CC   sequence and the entire, partial or derivative of purF gene, is used for
CC   mass production of the antimicrobial peptide in microorganisms without
CC   killing the host cells. Use of the purF gene derivative sequence,
CC   neutralises the toxicity of the antimicrobial peptides against the host
CC   microorganism. The antimicrobial peptides are useful commercially in the
CC   pharmaceutical and food industries.
XX
SQ   Sequence 53 BP; 8 A; 12 C; 20 G; 13 T; 0 other;


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAZ29364  ..

 Align seg 1/1  to: AAZ29364  from: 1  to: 53

       4 LysTrpProTrpTrpProTrpArgArg 12
         |||||||||||||||||||||||||||
      17 AAATGGCCGTGGTGGCCGTGGCGTCGT 43

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ40246

seq_documentation_block:
ID   AAZ40246 standard; DNA; 69 BP.
XX
AC   AAZ40246;
XX
DT   23-FEB-2000  (first entry)
XX
DE   Oligonucleotide for cloning indolicidin peptide coding sequence.
XX
KW   Indolicidin; bactenecin; sulphate-reducing bacteria; growth inhibitor;
KW   corrosion; degradation; metal; concrete; cement; dental implant; biofilm;
KW   ss.
XX
OS   Synthetic.
OS   Bacillus sp.
XX
PN   WO9956553-A1.
XX
PD   11-NOV-1999.
XX
PF   03-MAY-1999;   99WO-US09675.
XX
PR   06-MAY-1998;   98US-0074037.
PR   31-MAR-1999;   99US-0282277.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Wood TK,  Jayaraman A,  Earthman JC;
XX
DR   WPI; 2000-052882/04.
XX
PT   Inhibiting growth of sulphate-reducing bacteria using other bacteria,
PT   particularly for protection of metals and concrete  -
XX
```

```
PS   Example 4; Fig 1; 84pp; English.
XX
CC   This sequence represents an oligonucleotide for cloning the non-amidated
CC   indolicidin peptide coding sequence. The invention relates to a method
CC   for inhibiting growth of sulphate-reducing bacteria (A) on a material (B)
CC   sensitive to corrosion or degradation, by applying to (B) a bacterium (C)
CC   that secretes a compound (I) able to inhibit growth of (A). The method is
CC   used to protect metal, concrete or cement against corrosion and
CC   degradation, but (B) can also be used to protect dental implants. (B) is
CC   present in an open or closed system (e.g. water cooling tower, liquid
CC   storage container, fuel tank, sewer or drainage system etc.) or part of a
CC   bridge or other structure. The method is more effective and less
CC   expensive than known methods for inhibiting (A), and reduces the amount
CC   of toxic chemicals released. Conventional biofilms of aerobic organisms
CC   tend to encourage growth of (A), and addition of (C) to the biofilm
CC   prevents this. A single application of (C) lasts for a long time, and (I)
CC   are produced exactly where they are required and inhibit (A) without
CC   significant impact on other organisms (this effect includes reducing
CC   resistance of (A) to conventional biocides, which may then be used in
CC   reduced amounts). If local damage to the biofilm occurs, the underlying
CC   material is still protected by diffusion of (I) from neighbouring areas.
XX
SQ   Sequence 69 BP; 14 A; 18 C; 20 G; 17 T; 0 other;


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAZ40246   ..

 Align seg 1/1  to: AAZ40246  from: 1  to: 69

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||||
      28 AAATGGCCTTGGTGGCCTTGGCGCCGC 54

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ49764

seq_documentation_block:
ID   AAZ49764 standard; DNA; 211 BP.
XX
AC   AAZ49764;
XX
DT   18-APR-2000  (first entry)
XX
DE   Poly-(Indol (1-13)-Met-Ala-Arg-Ile-Ala-Met)3 DNA.
XX
KW   Crosslinked indolicidin analog; X-indolicidin; poly-Indol 1-13;
KW   stability; bovine neutrophil; antimicrobial; antibacterial; fungicide;
KW   protozoacide; virucide; anti-HIV; human immunodeficiency virus-1;
KW   HIV-1; gram positive bacteria; gram negative; Staphylococcus aureus;
KW   Escherichia coli; Salmonella typhimurium; yeast; fungi; protozoa;
KW   Candida albicans; Cryptococcus neoformans; Giardia; Acanthamoeba;
KW   hexapeptide spacer; ds..
XX
OS   Synthetic.
OS   Bos sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             8..199
FT                   /*tag=  a
FT                   /product= "Poly-(Indol(1-13)-Met-Ala-Arg-Ile-Ala-Met)3"
FT                   /note= "encodes three copies of Indol 1-13, each
FT                   separated by Met-Ala-Arg-Ile-Ala-Met spacer sequence"
FT   primer_bind     1..21
FT                   /*tag=  b
FT   primer_bind     complement (191..211)
FT                   /*tag=  c
FT   misc_feature    68..71
FT                   /*tag=  d
FT                   /note= "corresponds to overlap in oligonucleotides
FT                   used for ligation"
FT   misc_feature    148..151
FT                   /*tag=  e
FT                   /note= "corresponds to overlap in oligonucleotides
FT                   used for ligation"
XX
PN   WO9965510-A1.
XX
PD   23-DEC-1999.
XX
PF   20-MAY-1999;   99WO-US11165.
XX
PR   18-JUN-1998;   98US-0099631.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Selsted ME,  Osapay K;
XX
DR   WPI; 2000-147133/13.
DR   P-PSDB; AAY44668.
XX
PT   Crosslinked indolicidin analogs with antimicrobial activity against
PT   bacteria, yeast, fungi, protozoa and viruses  -
XX
PS   Example 1C; Fig 1; 53pp; English.
XX
CC   The patent discloses crosslinked analogs of indolicidin (Indol 1-13)
CC   which is a naturally occurring peptide isolated from bovine neutrophils
CC   and has antimicrobial activity. The crosslinked indolicidin
CC   (X-indolicidin) analogs are stable and have antimicrobial activity
CC   against gram positive and negative bacteria (e.g. Staphylococcus aureus,
CC   Escherichia coli and Salmonella typhimurium), yeasts and fungi (e.g.
CC   Candida albicans, Cryptococcus neoformans), protozoa (e.g. Giardia
CC   species and Acanthamoeba species), and viruses (e.g. HIV-1).
CC   They can be used for reducing or inhibiting the growth or survival of
CC   microorganisms in an environment e.g. a food or food product, a
CC   solution, an inanimate object comprising a surface, or a mammal.
CC   The present sequence is a DNA encoding a protein comprising three
CC   copies of Indol 1-13 each separated by a hexapeptide spacer sequence.
CC   The sequence was used to produce a recombinant construct for the
CC   expression of Indol-homoserine (Hse) analog. The ability of
CC   Indol-Hse analog to maintain antimicrobial activity provides a means to
CC   produce X-indolicidin analog precursors in sufficient quantities.
XX
SQ   Sequence 211 BP; 36 A; 50 C; 74 G; 51 T; 0 other;


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAZ49764   ..

 Align seg 1/1  to: AAZ49764  from: 1  to: 211

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||||
      38 AAATGGCCGTGGTGGCCGTGGCGTCGT 64

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ45123

seq_documentation_block:
ID   AAZ45123 standard; DNA; 211 BP.
XX
AC   AAZ45123;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Indolicidin fusion peptide nucleotide sequence.
XX
```

KW Indolicidin analogue; antimicrobial activity; helminth; bacteria; virus;
KW treatment; inhibit growth; micro-organism; contact lens solution;
KW transgenic plant; surgical instrument; yeast; fungi; protozoa; ss.
XX
OS Synthetic.
XX
PN WO9958141-A1.
XX
PD 18-NOV-1999.
XX
PF 05-MAY-1999; 99WO-US09942.
XX
PR 12-MAY-1998; 98US-0076227.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Selsted ME;
XX
DR WPI; 2000-053028/04.
DR P-PSDB; AAY57142.
XX
PT New indolicidin analogues, active against bacteria, yeast, fungi,
PT protozoa and virus, used for, e.g. treating infections -
XX
PS Disclosure; Fig 6; 62pp; English.
XX
CC This is the nucleotide sequence of an example of a fusion protein which
CC consists of an indolicidin analogue linked to another peptide.
CC Peptides AAY57109-Y57138 and AAY57143-Y57144 are new indolicidin
CC analogues, which have a homoserine residue and/or a truncated amino
CC terminal region. The analogues have the following amino acid sequence:
CC Xaa1-Xaa2-Xaa3-Xaa4-Xaa5-Xaa6-Pro-Xaa6-Xaa6-Pro-Xaa6-Xaa7-Xaa7-Xaa8
CC Where:
CC Xaa1 = Ile, Leu, Val, Ala, Gly or absent;
CC Xaa2 = Ile, Leu, Val, Ala, Gly or absent;
CC Xaa3 = Pro or absent;
CC Xaa4 = Trp, Phe or absent;
CC Xaa5 = Arg, Lys or absent;
CC Xaa6 = Trp or Phe;
CC Xaa7 = Arg, Lys or absent;
CC Xaa8 = homoserine (Hse), Met, Met-Xaa9-Met or absent, and
CC Xaa9 = at least one amino acid;
CC provided that if Xaa1 is present, Xaa8 = Hse, Met or Met-Xaa9-Met;
CC and further provided that: If Xaa2 is absent, Xaa1 is absent; if Xaa3 is
CC absent, Xaa1 and Xaa2 are absent; if Xaa4 is absent, Xaa1, Xaa2 and Xaa3
CC are absent; and if Xaa5 is absent, Xaa1, Xaa2, Xaa3 and Xaa4 are absent.
CC The indolicidin analogues can be used to create a fusion polypeptide
CC consisting of the analogue linked to a peptide. The indolicidin
CC analogues have antimicrobial activity against gram positive bacteria,
CC gram negative bacteria, yeast, fungus, protozoa and viruses (e.g. HIV-1).
CC They are also active against helminths. The analogues can be used for
CC reducing or inhibiting growth or survival of a microorganism. They can be
CC used for treating infections. They can also be included in a liquid such
CC as water or an aqueous solution, e.g. contact lens solution. The
CC analogues have potential uses in food products, and in objects such as
CC the surface of an instrument used to prepare food or to perform surgery.
CC Transgenic plants or animals useful in the food industry can be produced
CC by introducing a nucleic acid molecule encoding an indolicidin analogue
CC into the germline cells of such organisms.
XX
SQ Sequence 211 BP; 36 A; 50 C; 74 G; 51 T; 0 other;

alignment_scores:
Quality: 73.00 Length: 9
Ratio: 8.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-444-281-35 x AAZ45123 ..

Align seg 1/1 to: AAZ45123 from: 1 to: 211

4 LysTrpProTrpTrpProTrpArgArg 12
||||||||||||||||||||||||||
38 AAATGGCCGTGGTGGCCGTGGCGTCGT 64

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ20646

seq_documentation_block:
ID AAZ20646 standard; RNA; 6446 BP.
XX
AC AAZ20646;
XX
DT 26-NOV-1999 (first entry)
XX
DE TMV-based virus TMV861 coat protein read-through RNA sequence.
XX
KW TMV-based virus; tobacco mosaic virus; protein isolation; green juice;
KW virus isolation; fusion protein identification; ss.
XX
OS Tobacco mosaic virus.
XX
PN WO9946288-A2.
XX
PD 16-SEP-1999.
XX
PF 09-MAR-1999; 99WO-US05056.
XX
PR 10-MAR-1998; 98US-0037751.
XX
PA (BIOS-) BIOSOURCE TECHNOLOGIES INC.
XX
PI Garger SJ, Holtz RB, McCulloch MJ, Turpen TH;
XX
DR WPI; 1999-561660/47.
XX
PT Obtaining protein, viruses and fusion proteins from plants, using
PT non-denaturing conditions -
XX
PS Disclosure; Page 55-58; 58pp; English.
XX
CC This sequence represents a tobacco mosaic virus (TMV) based virus
CC sequence identified using the method of the invention. The method is for
CC obtaining a soluble protein or peptide of interest from a plant,
CC comprises homogenising the plant to produce green juice, adjusting the
CC pH to less than or equal to 5.2, and heating the juice to a minimum of
CC 45 degrees C. The juice is then centrifuged to produce a supernatant, and
CC the protein or peptide is purified from the supernatant. The method can
CC also be used for obtaining viruses and fusion proteins. The method is
CC especially useful for obtaining IL-1 to IL-10, EPO, G-CSF, GM-CSF,
CC hP-CSF, M-CSF, Factor VIII, Factor IX, tPA, receptors, receptor
CC antagonists, antibodies, single-chain antibodies, enzymes,
CC neuropolypeptides, insulin, antigens, vaccines, peptide hormones,
CC calcitonin, and human growth hormone, or an antimicrobial peptide or
CC protein from protegrins, magainins, cecropins, melittins, indolicidins,
CC defensins, beta-defensins, cryptdins, clavainins, plant defensins,
CC nicin and bactenecins, all produced by recombinant means. The new method
CC is more efficient than the prior art for isolating viruses, protein, and
CC peptides. The method is large-scale, and non-denaturing and
CC solvent-limited. Prior art methods do not isolate recombinant proteins,
CC and do not allow fraction 2 proteins to be ultrafiltrated.
XX
SQ Sequence 6446 BP; 1873 A; 1234 C; 1563 G; 1776 U; 0 other;

alignment_scores:
Quality: 73.00 Length: 9
Ratio: 8.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-444-281-35 x AAZ20646 ..

Align seg 1/1 to: AAZ20646 from: 1 to: 6446

```
      4 LysTrpProTrpTrpProTrpArgArg 12
        ||||||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF82334

seq_documentation_block:
ID   AAF82334 standard; RNA; 6446 BP.
XX
AC   AAF82334;
XX
DT   22-JUN-2001  (first entry)
XX
DE   Tobacco mosaic virus-based coat protein read-through virus TMV861.
XX
KW   TMV; tobacco mosaic virus; TMV861; virus isolation;
KW   non-native protein purification; ribulose 1,5-diphosphate carboxylase;
KW   RuBisCo; coat protein read-through; ss.
XX
OS   Tobacco mosaic virus.
XX
PN   WO200119969-A1.
XX
PD   22-MAR-2001.
XX
PF   19-MAY-2000; 2000WO-US13680.
XX
PR   16-SEP-1999;   99US-0397090.
XX
PA   (LARG-) LARGE SCALE BIOLOGY CORP.
XX
PI   Garger SJ,  Holtz BR,  McCulloch MJ,  Turpen TH;
XX
DR   WPI; 2001-328016/34.
XX
PT   Minimizing presence of ribulose 1,5-diphosphate carboxylase to obtain
PT   plant product for isolating bioactive species involves cutting plant
PT   material from plant in cutting period when quantity of RuBisCo is at
PT   minimum  -
XX
PS   Disclosure; Page 72-74; 81pp; English.
XX
CC   The present sequence is a tobacco mosaic virus (TMV)-based virus
CC   which was used to infect field-grown tobacco. The virus was then
CC   isolated from the tobacco plants by a novel process for isolating and
CC   purifying viruses, soluble proteins and peptides from plant sources. In
CC   order to isolate the bioactive species from the undesirable
CC   photosynthetic protein ribulose 1,5-diphosphate carboxylase (RuBisCo),
CC   the plant material is cut in a period of the light/dark cycle when the
CC   quantity of RuBisCo in the plant is at a minimum. The method is useful
CC   for obtaining a virus of interest. It is also useful for obtaining
CC   soluble recombinant or non-native proteins, such as active mammalian
CC   proteins, enzymes, vaccines, antibodies and peptides, from transgenic
CC   plants.
XX
SQ   Sequence 6446 BP; 1873 A; 1234 C; 1563 G; 1776 U; 0 other;


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AAF82334   ..

 Align seg 1/1  to: AAF82334  from: 1  to: 6446

      4 LysTrpProTrpTrpProTrpArgArg 12
        ||||||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28519

seq_documentation_block:
ID   AAA28519 standard; DNA; 207 BP.
XX
AC   AAA28519;
XX
DT   29-AUG-2000  (first entry)
XX
DE   PCRIL DNA coding sequence.
XX
KW   Magainin; antimicrobial; transgenic plant; protease degradation; Rev4;
KW   indolicidin; protein production; reverse peptide; ss.
XX
OS   Synthetic.
XX
PN   WO200026344-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;   99WO-US25561.
XX
PR   30-OCT-1998;   98US-0106373.
PR   02-NOV-1998;   98US-0106537.
XX
PA   (INTE-) INTERLINK BIOTECHNOLOGIES LLC.
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Everett NP,  Li Q,  Lawrence C,  Davies MH;
XX
DR   WPI; 2000-365597/31.
DR   P-PSDB; AAY92840.
XX
PT   Polypeptides for reducing proteolytic degradation of proteins
PT   administered to, or produced by a plant comprise indolicin or its
PT   functional equivalents
XX
PS   Example 17; Page 35; 50pp; English.
XX
CC   Indolicidin is a potent antimicrobial tridecapeptide, originally
CC   purified from cytoplasmic granules of bovine neutrophils.  Reverse
CC   peptide, Rev4 of indolicidin (see AAY92794) was found to have increased
CC   stability against plant protease degradation. Expression of antimicrobial
CC   peptides in transgenic plants suffers a major limitation in that the
CC   foreign peptides are susceptible to rapid degradation by proteases. The
CC   invention concerns reducing the extent of protease degradation of a
CC   protein applied to, or produced by a plant by administering indolicidin,
CC   Rev4 or a functional equivalent to the plant. Transgenic plants
CC   expressing indolicidin and Rev4 are useful for production of the
CC   antimicrobial peptides. Compositions containing indolicidin and Rev4 are
CC   also useful for production of agronomically important proteins in
CC   plants.
XX
SQ   Sequence 207 BP; 49 A; 50 C; 36 G; 72 T; 0 other;


alignment_scores:
            Quality:   66.00              Length:     10
              Ratio:  6.600                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 60.000

alignment_block:
 US-09-444-281-35 x AAA28519   ..

 Align seg 1/1  to: AAA28519  from: 1  to: 207

      2 LeuLysLysTrpProTrpTrpProTrpArg 11
        :::::::::||||||||||||||||||:::
    163 ATTAGGAGATGGCCTTGGTGGCCTTGGAAA 192
```

```
OM of: US-09-444-281-35 to: Issued_Patents_NA:*  out_format : pfs

Date: Jan 4, 2002 11:03 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084144_16216/app_query.fasta_1.210
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09444281_@CGN1_1_70 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-35
Query length: 13
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 135.410000


score_list:
Sequence          Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-071-49 +   91.00  202.85  0.0012    88  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-230-180-29 +   73.00  172.79  0.0565    39  !  S
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-259-741-5 +   73.00  131.92  10.69     6446  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-037-751-5 +   73.00  131.92  10.69     6446  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-466-422-5 +   73.00  131.92  10.69     6446  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-159-784-1 -   60.00  109.27  195.21    4031  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-337A-1 +   58.00  113.07  119.89    1509  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-696-268B-1 +   58.00  113.07  119.89    1509  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05467-1 +   58.00  113.07  119.89    1509
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-11328A-3 +   58.00  112.62  126.99    1596
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-382-256-1 +   58.00  110.88  158.78    1984  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-395-115-1 +   58.00  110.88  158.78    1984  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-816-1 +   57.50  110.24  172.45    1894  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-149-097D-24 +   57.50   99.74  662.83    7032  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-949-386-24 +   57.50   99.74  662.83    7032  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-450-562-24 +   57.50   99.74  662.83    7032  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-949-386-25 +   57.50   99.67  668.35    7089  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-450-562-25 +   57.50   99.67  668.35    7089  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-3 +   56.00  115.35  89.48      683  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-1 +   56.00  111.47  147.16    1109  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-6 +   56.00  111.47  147.16    1109  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2 +   55.00   43.12  5.2e+05   4403765
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-920-812-13 +   54.00   90.21  2.3e+03   9515  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-920-827-13 +   54.00   90.21  2.3e+03   9515  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-921-177-13 +   54.00   90.21  2.3e+03   9515  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-362-577C-13 +   54.00   90.21  2.3e+03   9515  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-920-828-13 +   54.00   90.21  2.3e+03   9515  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-153-804-10 +   53.50  103.08  431.52    1677  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-438-439C-1 +   53.00  105.09  333.74    1150  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-705-245-19 +   53.00  105.09  333.74    1150  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-06665-1 +   53.00  105.09  333.74    1150
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-234-332-6 +   53.00   98.49  777.98    2623  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-417-1 +   53.00   91.42  1.9e+03   6344  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-905-223-108 +   52.50  115.43  88.52      278  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-057-860A-1 +   52.50  109.16  197.98     609  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-026-958-1 +   52.50  101.15  552.75    1656  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-808-931-14 -   52.50   94.41  1.3e+03   3848  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-808-323-14 -   52.50   94.41  1.3e+03   3848  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-050-603A-14 -   52.50   94.41  1.3e+03   3848  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-420B-14 -   52.50   94.41  1.3e+03   3848  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-020-956-82 +   52.00  111.85  140.10     383
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-82 +   52.00  111.85  140.10     383
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-183 +   52.00  111.83  140.48     384
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-020-956-73 +   52.00  109.74  183.82     499
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-73 +   52.00  109.74  183.82     499

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-071-49

seq_documentation_block:
; Sequence 49, Application US/09042071
; Patent No. 6294372
;  GENERAL INFORMATION:
;    APPLICANT:  Burian, Jan
;    APPLICANT:  Kay, William W.
;    TITLE OF INVENTION:  REPLICATION GENES AND GENE PRODUCTS FROM
;    TITLE OF INVENTION:  SMALL CRYPTIC PLASMIDS AND METHODS FOR CONSTRUCTING
;    TITLE OF INVENTION:  CONTROLLED-REPLICATION PLASMID VECTORS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,071
;      FILING DATE:  13-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  660081.407
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  88 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-042-071-49


alignment_scores:
            Quality:   91.00              Length:     13
              Ratio:  7.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-042-071-49   ..

 Align seg 1/1  to: US-09-042-071-49  from: 1  to: 88

        1 IleLeuLysLysTrpProTrpTrpProTrpArgArgLys 13
          |||||||||||||||||||||||||||||||||||||||
       25 ATCCTGAAAAAATGGCCGTGGTGGCCGTGGCGTCGTAAA 63

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-230-180-29

seq_documentation_block:
; Sequence 29, Application US/09230180
; Patent No. 6183992
; GENERAL INFORMATION:
;  APPLICANT:  Kim, Sun-Chang
;  APPLICANT:  Lee, Jae Hyun
;  APPLICANT:  Kang, Min Hyung
```

```
;  APPLICANT:  Kim, Jeong Hyun
;  APPLICANT:  Hong, Seung-Suh
;  APPLICANT:  Lee, Hyun-Soo
;  APPLICANT:  Samyang Genex Corporation
;  APPLICANT:  Korea Advanced Institute of Science and Technology
;  TITLE OF INVENTION:  METHOD FOR MASS PRODUCTION OF
;  TITLE OF INVENTION:  ANTIMICROBIAL PEPTIDE
;  FILE REFERENCE:  6181/0F135
;  CURRENT APPLICATION NUMBER: US/09/230,180
;  CURRENT FILING DATE:  1999-03-10
;  PRIOR APPLICATION NUMBER:  PCT/KR98/00132
;  PRIOR FILING DATE:  1998-05-28
;  PRIOR APPLICATION NUMBER:  KR 13372/1998
;  PRIOR FILING DATE:  1998-04-09
;  PRIOR APPLICATION NUMBER:  KR 21312/1997
;  PRIOR FILING DATE:  1997-05-28
;  NUMBER OF SEQ ID NOS:  36
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 39
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DNA sequence deduced from Indolicidin peptide
;   OTHER INFORMATION: sequence based on codon usage of E. coli
US-09-230-180-29


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-230-180-29   ..

 Align seg 1/1  to: US-09-230-180-29  from: 1  to: 39

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||
      13 AAATGGCCGTGGTGGCCGTGGCGTCGT 39

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-259-741-5

seq_documentation_block:
; Sequence 5, Application US/09259741
; Patent No. 6033895
;  GENERAL INFORMATION:
;    APPLICANT:  GARGER, STEPHEN
;    APPLICANT:  HOLTZ, R. BARRY
;    APPLICANT:  McCULLOCH, MICHAEL
;    APPLICANT:  TURPEN, THOMAS
;    TITLE OF INVENTION:  A PROCESS FOR ISOLATING AND
;    TITLE OF INVENTION:  PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
;    TITLE OF INVENTION:  SOURCES
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howrey & Simon
;      STREET:  1299 Pennsylvania Avenue N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/259,741
;      FILING DATE:  February 25, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/037,751
;      FILING DATE:  March 10, 1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P
;      REGISTRATION NUMBER:  25,277
;      REFERENCE/DOCKET NUMBER:  00801.0140.US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-463-8100
;      TELEFAX:  650-463-8400
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6446 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  Genomic RNA
US-09-259-741-5


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-259-741-5   ..

 Align seg 1/1  to: US-09-259-741-5  from: 1  to: 6446

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-037-751-5

seq_documentation_block:
; Sequence 5, Application US/09037751
; Patent No. 6037456
;  GENERAL INFORMATION:
;    APPLICANT:  GARGER, STEPHEN
;    APPLICANT:  HOLTZ, R. BARRY
;    APPLICANT:  McCULLOCH, MICHAEL
;    APPLICANT:  TURPEN, THOMAS
;    TITLE OF INVENTION:  A PROCESS FOR ISOLATING AND
;    TITLE OF INVENTION:  PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
;    TITLE OF INVENTION:  FROM PLANT SOURCES
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howrey & Simon
;      STREET:  1299 Pennsylvania Avenue N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,751
;      FILING DATE:  10-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P
;      REGISTRATION NUMBER:  25,277
;      REFERENCE/DOCKET NUMBER:  00801.0140.999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-463-8109
```

```
;       TELEFAX:  650-463-8400
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6446 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  Genomic RNA
US-09-037-751-5


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-037-751-5   ..

 Align seg 1/1  to: US-09-037-751-5  from: 1  to: 6446

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-466-422-5

seq_documentation_block:
; Sequence 5, Application US/09466422
; Patent No. 6303779
;   GENERAL INFORMATION:
;       APPLICANT: GARGER, STEPHEN
;                  HOLTZ, R. BARRY
;                  McCULLOCH, MICHAEL
;                  TURPEN, THOMAS
;       TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
;                           PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
;                           FROM PLANT SOURCES
;       NUMBER OF SEQUENCES: 5
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howrey & Simon
;            STREET: 1299 Pennsylvania Avenue N.W.
;            CITY: Washington
;            STATE: DC
;            COUNTRY: USA
;            ZIP: 20004
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/466,422
;            FILING DATE: 17-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/037,751
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Halluin, Albert P
;            REGISTRATION NUMBER: 25,277
;            REFERENCE/DOCKET NUMBER: 00801.0140.999
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650-463-8109
;            TELEFAX: 650-463-8400
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 6446 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: unknown
;       MOLECULE TYPE: Genomic RNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-466-422-5


alignment_scores:
            Quality:   73.00              Length:      9
              Ratio:  8.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-466-422-5   ..

 Align seg 1/1  to: US-09-466-422-5  from: 1  to: 6446

       4 LysTrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-159-784-1

seq_documentation_block:
; Sequence 1, Application US/08159784
; Patent No. 5643783
;  GENERAL INFORMATION:
;    APPLICANT:  Bjorn R. Olsen
;    TITLE OF INVENTION:  NOVEL COLLAGEN AND USES THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/159,784
;      FILING DATE:  December 1, 1993
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John F. Freeman
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00246/170001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4031
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-159-784-1


alignment_scores:
            Quality:   60.00              Length:      9
              Ratio:  7.500                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-444-281-35 x US-08-159-784-1/rev   ..

 Align seg 1/1  to reverse of: US-08-159-784-1  from: 1  to: 4031
```

```
      3 LysLysTrpProTrpTrpProTrpArg 11
        |||:::|||   ||||||||||||||
   1671 AAATCCTGGTCCTGGTGGCCCTGGAGG 1645

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-337A-1

seq_documentation_block:
; Sequence 1, Application US/08481337A
; Patent No. 5863738
;  GENERAL INFORMATION:
;    APPLICANT:  TEN DIJKE, Peter
;    APPLICANT:  HELDIN, Carl-Henrik
;    APPLICANT:  MIYAZONO, Kohei
;    APPLICANT:  SAMPATH, Kuber T.
;    TITLE OF INVENTION:  Morphogenic Protein-Specific Cell
;    TITLE OF INVENTION:  Surface Receptors and Uses Therefor
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,337A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  CRP-097CP2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1509
;      OTHER INFORMATION:  /product= "Human ALK1"
US-08-481-337A-1


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-08-481-337A-1    ..

 Align seg 1/1  to: US-08-481-337A-1  from: 1  to: 1509

       5 TrpProTrpTrpProTrp 10
         ||||||||||||||||||
     389 TGGCCCTGGTGGCCCTGG 406

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-696-268B-1

seq_documentation_block:
; Sequence 1, Application US/08696268B
; Patent No. 5968752
;  GENERAL INFORMATION:
;    APPLICANT:  ICHIJO, HIDENORI
;    APPLICANT:  NISHITOH, HIDEKI
;    APPLICANT:  SAMPATH, KUBER T.
;    TITLE OF INVENTION:  NOVEL SIGNALLING RECEPTOR FOR
;    TITLE OF INVENTION:  MORPHOGENIC PROTEINS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,268B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  CRP-117
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1509
;      OTHER INFORMATION:  /product= "Human ALK-1"
US-08-696-268B-1


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-08-696-268B-1    ..

 Align seg 1/1  to: US-08-696-268B-1  from: 1  to: 1509

       5 TrpProTrpTrpProTrp 10
         ||||||||||||||||||
     389 TGGCCCTGGTGGCCCTGG 406

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05467-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9505467
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN-SPECIFIC CELL
;    TITLE OF INVENTION:  SURFACE RECEPTORS AND USES THEREFOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, TESTA, HURWITZ &
;      ADDRESSEE:  THIBEAULT
;      STREET:  53 STATE STREET
```

```
;       CITY:  BOSTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/05467
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-097PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7000
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1509 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1509
;       OTHER INFORMATION:  /product= "Human ALK1"
PCT-US95-05467-1


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x PCT-US95-05467-1   ..

 Align seg 1/1  to: PCT-US95-05467-1  from: 1  to: 1509

       5 TrpProTrpTrpProTrp 10
         ||||||||||||||||||
     389 TGGCCCTGGTGGCCCTGG 406

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-11328A-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9411328A
;  GENERAL INFORMATION:
;     APPLICANT:  HE, ET AL.
;     TITLE OF INVENTION:  TAR-1 and TAR-3
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;       ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;       STREET:  6 BECKER FARM ROAD
;       CITY:  ROSELAND
;       STATE:  NEW JERSEY
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 INCH DISKETTE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/11328A
;       FILING DATE:  Submitted herewith
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FERRARO, GREGORY D.
;       REGISTRATION NUMBER:  36,134
;       REFERENCE/DOCKET NUMBER:  325800-132
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1596 BASE PAIRS
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  SINGLE
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  cDNA
PCT-US94-11328A-3


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x PCT-US94-11328A-3   ..

 Align seg 1/1  to: PCT-US94-11328A-3  from: 1  to: 1596

       5 TrpProTrpTrpProTrp 10
         ||||||||||||||||||
     470 TGGCCCTGGTGGCCCTGG 487

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-382-256-1

seq_documentation_block:
; Sequence 1, Application US/09382256A
; Patent No. 6207814
;   GENERAL INFORMATION:
;       APPLICANT: MIYAZONO, Kohei
;                  TEN DIJKE, Peter
;                  FRANZEN, Petra
;                  YAMASHITA, Hidetoshi
;                  HELDIN, Carl-Henrik
;       TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS
;                           HAVING SERINE THREONINE KINASE DOMAINS,
;                           AND THEIR USE
;       NUMBER OF SEQUENCES: 29
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fulbright & Jaworski L.L.P.
;            STREET: 666 Fifth Avenue
;            CITY: New York City
;            STATE: New York
;            COUNTRY: USA
;            ZIP: 10103
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
;            COMPUTER: IBM PS/2
;            OPERATING SYSTEM: PC-DOS
;            SOFTWARE: Wordperfect
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/382,256A
;            FILING DATE: 24-Aug-1999
;            CLASSIFICATION: 514
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB93/02367
;            FILING DATE: No. 6207814ember 17, 1993
;            APPLICATION NUMBER: GB 9224057.1
;            FILING DATE: No. 6207814ember 17, 1992
;            APPLICATION NUMBER: GB 9304677.9
;            FILING DATE: March 8, 1993
;            APPLICATION NUMBER: GB 9304680.3
```

```
;             FILING DATE: March 8, 1993
;             APPLICATION NUMBER: 9311047.6
;             FILING DATE: May 28, 1993
;             APPLICATION NUMBER: 9313763.6
;             FILING DATE: July 2, 1993
;             APPLICATION NUMBER: 9316099.2
;             FILING DATE: August 3, 1993
;             APPLICATION NUMBER: 321344.5
;             FILING DATE: October 15, 1993
;       ATTORNEY/AGENT INFORMATION:
;             NAME: No. 6207814man D. Hanson
;             REGISTRATION NUMBER: 30,946
;             REFERENCE/DOCKET NUMBER: LUD 5298.1
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (212) 318-3000
;             TELEFAX: (212) 752-5958
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 1984 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: unknown
;             TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       FRAGMENT TYPE: internal
;       ORIGINAL SOURCE:
;             ORGANISM: Homo sapiens
;       FEATURE:
;             NAME/KEY: CDS
;             LOCATION: 283..1791
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-382-256-1


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-382-256-1   ..

 Align seg 1/1  to: US-09-382-256-1  from: 1  to: 1984

       5 TrpProTrpTrpProTrp 10
         ||||||||||||||||||
     671 TGGCCCTGGTGGCCCTGG 688

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-395-115-1

seq_documentation_block:
; Sequence 1, Application US/09395115
; Patent No. 6271365
;  GENERAL INFORMATION:
;    APPLICANT:  Miyazono, Kohei; Dijke, Peter Ten;
;    APPLICANT:  Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
;    TITLE OF INVENTION:  Activin Receptor-Like Kinase, Proteins
;    TITLE OF INVENTION:  Having Serine Threonine Kinase Domains And Their Use
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/395,115
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/436,265
;      FILING DATE:  30-October-1995
;      APPLICATION NUMBER:  PCT/GB93/02367
;      FILING DATE:  17-No. 6271365ember-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9224057.1
;      FILING DATE:  17-No. 6271365ember-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9304677.9
;      FILING DATE:  8-March-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9304680.3
;      FILING DATE:  8-March-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9311047.6
;      FILING DATE:  28-May-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9313763.6
;      FILING DATE:  2-July-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9136099.2
;      FILING DATE:  3-August-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9321344.5
;      FILING DATE:  15-October-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohlei, Vineet
;      REGISTRATION NUMBER:  37,003
;      REFERENCE/DOCKET NUMBER:  LUD 5298
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1984 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  283..1791
US-09-395-115-1


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x US-09-395-115-1   ..

 Align seg 1/1  to: US-09-395-115-1  from: 1  to: 1984

       5 TrpProTrpTrpProTrp 10
         ||||||||||||||||||
     671 TGGCCCTGGTGGCCCTGG 688

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-816-1

seq_documentation_block:
; Sequence 1, Application US/08749816
; Patent No. 6013470
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Lesage, Florian
;    APPLICANT:  Guillemare, Eric
;    APPLICANT:  Fink, Michel
;    APPLICANT:  Duprat, Fabrice
;    APPLICANT:  Lazdunki, Michel
;    APPLICANT:  Romey, Georges
;    APPLICANT:  Barhanin, Jacques
;    TITLE OF INVENTION:  FAMILY OF MAMMALIAN POTASSIUM CHANNELS,
;    TITLE OF INVENTION:  THEIR CLONING AND THEIR USE ESPECIALLY FOR THE SCREENING
;    TITLE OF INVENTION:  OF DRUGS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WEISER & ASSOCIATES
;      STREET:  230 South Fifteenth Street, Suite 500
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/749,816
;      FILING DATE:  15-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weiser, Gerard J.
;      REGISTRATION NUMBER:  39,763
;      REFERENCE/DOCKET NUMBER:  989.6351P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-875-8883
;      TELEFAX:  215-875-8394
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..1190
US-08-749-816-1


alignment_scores:
            Quality:   57.50              Length:     13
              Ratio:  5.227                Gaps:      1
 Percent Similarity:  84.615   Percent Identity: 61.538

alignment_block:
 US-09-444-281-35 x US-08-749-816-1    ..

 Align seg 1/1  to: US-08-749-816-1  from: 1  to: 1894

       2 LeuLysLysTrpProTrp...TrpProTrpArgArgLys 13
         |||::::::|||||||||   |||   ||||:||||:::
     136 CTGCGGCGTTGGCCTTGGCTTTGGCTTTGGCGGCGGCGG 174

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-149-097D-24

seq_documentation_block:
; Sequence 24, Application US/08149097D
; Patent No. 5874236
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  Feldman, Daniel
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,097D
;      FILING DATE:  05-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,536
;      FILING DATE:  11-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US92/06903
;      FILING DATE:  14-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/914,231
;      FILING DATE:  13-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/868,354
;      FILING DATE:  10-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/745,206
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/620,250
;      FILING DATE:  30-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/482,384
;      FILING DATE:  20-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,751
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US89/01408
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/176,899
;      FILING DATE:  04-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6362-55038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 238-0999
;      TELEFAX:  (619) 238-0062
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  169..6921
;      OTHER INFORMATION:  /product= "Alpha1E-1 subunit of
;      OTHER INFORMATION:  human calcium channel"
US-08-149-097D-24


alignment_scores:
            Quality:   57.50              Length:     11
```

```
            Ratio:  6.389                Gaps:      1
 Percent Similarity:  81.818   Percent Identity: 72.727

alignment_block:
 US-09-444-281-35 x US-08-149-097D-24  ...

 Align seg 1/1  to: US-08-149-097D-24  from: 1  to: 7032

       2 LeuLysLysTrpProTrp...TrpProTrpArg 11
         |||:::   |||||||||   ||||||||||||
    2606 TTGAGGGCCTGGCCCTGGGCCTGGCCCTGGAGA 2638

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-949-386-24

seq_documentation_block:
; Sequence 24, Application US/08949386
; Patent No. 6090623
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Gillespie, Alison
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/949,386
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,012
;      FILING DATE:  11-AUG-1994
;      APPLICATION NUMBER:  08/149,097
;      FILING DATE:  5-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/105,536
;      FILING DATE:  11-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  519808
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 238-0999
;      TELEFAX:  (619) 238-0062
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7032 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  166..6921
;      OTHER INFORMATION:  /standard_name= "Alpha-1E-1"
 US-08-949-386-24


 alignment_scores:
            Quality:   57.50              Length:     11
```

```
            Ratio:  6.389                Gaps:      1
 Percent Similarity:  81.818   Percent Identity: 72.727

alignment_block:
 US-09-444-281-35 x US-08-949-386-24   ..

 Align seg 1/1  to: US-08-949-386-24  from: 1  to: 7032

       2 LeuLysLysTrpProTrp...TrpProTrpArg 11
         |||:::   |||||||||   ||||||||||||
    2606 TTGAGGGCCTGGCCCTGGGCCTGGCCCTGGAGA 2638
```

OM of: US-09-444-281-27 to: N_Geneseq_1101:* out_format : pfs

Date: Jan 4, 2002 11:02 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084144_16249/app_query.fasta_1.210
-DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09444281_@CGN1_1_267 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-444-281-27
Query length: 11
Database: N_Geneseq_1101:*
Database sequences: 930621
Database length: 428662619
Search time (sec): 286.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27297	+	54.00	181.09	0.0520	97	! C
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV90541	-	41.00	128.17	46.09	375	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH10202	+	41.00	124.81	70.95	557	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ12864	+	41.00	121.23	112.34	849	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28524	-	41.00	115.83	224.32	1601	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28526	-	41.00	112.79	331.24	2289	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28522	-	41.00	112.79	331.40	2290	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28525	-	41.00	112.58	340.56	2348	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28523	-	41.00	112.36	350.06	2408	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH17441	-	41.00	112.36	350.22	2409	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH18239	-	40.00	101.71	1.4e+03	5796	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH18247	-	40.00	101.35	1.4e+03	6048	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI61181	+	40.00	98.35	2.1e+03	8605	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI59395	-	40.00	98.12	2.2e+03	8840	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV68396	-	40.00	97.59	2.3e+03	9408	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX60096	-	40.00	92.52	4.5e+03	17091	
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF28542	+	40.00	80.91	2.0e+04	66986	
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ42995	-	39.00	121.36	110.45	395	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF11642	+	39.00	120.54	122.70	435	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA41342	+	39.00	119.39	142.19	498	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF93635	+	39.00	119.25	144.68	506	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF13345	-	39.00	116.60	203.22	691	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ24679	-	39.00	112.86	328.68	1074	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT29190	-	39.00	111.10	411.55	1320	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ24681	-	39.00	111.03	415.29	1331	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC77625	-	39.00	110.88	423.46	1355	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ29276	-	39.00	110.29	456.62	1452	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH98421	-	39.00	110.14	465.88	1479	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV48329	-	39.00	109.86	482.39	1527	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAS00585	-	39.00	109.62	497.91	1572	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:AAQ29277	-	39.00	108.91	545.40	1709	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD12577	-	39.00	108.71	559.33	1749	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV84372	-	39.00	107.39	662.94	2044	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC93390	-	39.00	107.15	683.12	2101	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF32713	-	39.00	107.15	683.12	2101	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA78392	+	39.00	105.16	882.40	2657	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD02761	+	39.00	101.28	1.5e+03	4194	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD02762	+	39.00	101.01	1.5e+03	4329	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD02760	+	39.00	100.83	1.5e+03	4422	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAD12951	+	39.00	99.58	1.8e+03	5122	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF83651	+	39.00	98.94	2.0e+03	5523	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF24892	+	39.00	87.83	8.1e+03	2039	!
/SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF11976	+	38.00	114.22	276.00	629	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:AAQ54498	+	38.00	112.49	344.58	771	!
/SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT02565	+	38.00	112.49	344.58	771	!

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA27297

seq_documentation_block:
ID AAA27297 standard; DNA; 97 BP.
XX
AC AAA27297;
XX
DT 20-SEP-2000 (first entry)
XX
DE Oligonucleotide used for synthesis of anionic spacer cassette.
XX
KW Oligonucleotide; cellulose binding domain; CBD; cationic peptide;
KW MBI-11; indolicidin; bovine; ss.
XX
OS Synthetic.
XX
PN WO200031279-A2.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-CA01107.
XX
PR 20-NOV-1998; 98US-0109218.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Burian J, Bartfeld D;
XX
DR WPI; 2000-400086/34.
XX
PT Multi-domain fusion protein expression cassette used for high yield
PT stable production of foreign peptide gene products -
XX
PS Example 5; Page 40; 73pp; English.
XX
CC A novel method allows the efficient production of cationic peptides in
CC recombinant host cells. The method involves construction of a
CC multi-domain fusion protein expression cassette comprising a promoter and
CC a nucleic acid molecule expressed as an insoluble protein. The inclusion
CC of anionic peptide sequences in the linker sequences neutralises the
CC positive charge of the cationic peptide so that the charge of the
CC fusion protein is controlled. This cassette allows high yield, stable
CC production of the cationic peptide. Cationic peptides such as
CC bovine indolicidin may be used as antimicrobial agents. The present
CC sequence is an oligonucleotide that was used to synthesise the
CC anionic spacer of the expression cassette.
XX
SQ Sequence 97 BP; 23 A; 29 C; 26 G; 19 T; 0 other;

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-444-281-27 x AAA27297 ..

Align seg 1/1 to: AAA27297 from: 1 to: 97

```
      1 HisGluAlaGluProGluAlaGluProIle 10
        ||||||||||||||||||||||||||||||
     38 CATGAAGCGGAACCGGAAGCGGAACCGATT 67
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV90541

seq_documentation_block:

```
ID   AAV90541 standard; cDNA; 375 BP.
XX
AC   AAV90541;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone BK517.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 574; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 375 BP; 56 A; 125 C; 113 G; 81 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:      9
              Ratio:  4.556                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAV90541/rev   ..

 Align seg 1/1  to reverse of: AAV90541  from: 1  to: 375

       1 HisGluAlaGluProGluAlaGluPro 9
         |||||||||||||||||:::::|||
     140 CACGAAGCAGAGCCTGAGGGGAGCCCG 114

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH10202

seq_documentation_block:
ID   AAH10202 standard; cDNA; 557 BP.
XX
AC   AAH10202;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:7037.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 7037; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 557 BP; 99 A; 165 C; 183 G; 103 T; 7 other;


alignment_scores:
            Quality:   41.00              Length:      9
              Ratio:  4.556                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAH10202   ..

 Align seg 1/1  to: AAH10202  from: 1  to: 557

       1 HisGluAlaGluProGluAlaGluPro 9
         |||||||||||||||||:::::|||
   253 CACGAAGCAGAGCCTGAGGGGAGCCCG 279
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ12864

seq_documentation_block:
ID   AAQ12864 standard; DNA; 849 BP.
XX
AC   AAQ12864;
XX
DT   11-OCT-1991  (first entry)
XX
DE   Human Cytotoxic Cell Protease-1 coding sequence.
XX
KW   hCCP1 inhibitor; cytotoxic T-lymphocytes; ss.
XX
OS   Homo sapiens.
XX
PN   WO9110685-A.
XX
PD   25-JUL-1991.
XX
PF   17-JAN-1991;   91WO-US00340.
XX
PR   19-JAN-1990;   90US-0467880.
XX
PA   (SERA-) SERAGEN INC.
XX
PI   Bleackley RC,  Lobe CG,  Paetkau VH,  James MN,  Murphy M;
XX
DR   WPI; 1991-237989/32.
XX
PT   DNA vectors, and inhibitors of cytotoxic cell protease - for
PT   treatment of auto-immune diseases e.g. pernicious anaemia,
PT   rheumatoid arthritis, allo-graft rejection etc.
XX
PS   Disclosure; Fig 6; 62pp; English.
XX
CC   The hCCP1 coding sequence was isolated from cytotoxic T-cell
CC   lymphocytes. Vectors comprising the hCCP1 coding sequence are
CC   claimed. Clone hC11 was isolated and found to be the human analogue
CC   of murine C11.
CC   See AAQ12862-6 and AAR13254-R13262.
XX
SQ   Sequence 849 BP; 232 A; 230 C; 222 G; 165 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:     11
              Ratio:  4.100                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x AAQ12864    ..

 Align seg 1/1  to: AAQ12864  from: 1  to: 849

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||||||||:::::::::|||   |||
      68 CATGAGGCCGAGCCCCACTCCCGCCCCTACATG 100

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28524

seq_documentation_block:
ID   AAA28524 standard; cDNA; 1601 BP.
XX
AC   AAA28524;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human opioid growth factor receptor cDNA spliced version 1.
XX
KW   OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW   cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..33
FT                   /*tag=  a
FT   CDS             34..1419
FT                   /*tag=  b
FT                   /product= Opioid_growth_factor_receptor
FT   3'UTR           1420..1601
FT                   /*tag=  c
XX
PN   WO200026340-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-US25802.
XX
PR   03-NOV-1998;   98US-0106879.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Zagon IS,  McLaughlin PJ,  Verderame MF;
XX
DR   WPI; 2000-365594/31.
DR   P-PSDB; AAY92810.
XX
PT   New cDNA encoding rat and human opioid growth factor receptors which
PT   modulate cell growth, useful for treating cancer
XX
PS   Claim 1; Page 82-83; 91pp; English.
XX
CC   Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC   used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC   human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC   consistently yielded a single species of cDNA, while the 3' RACE revealed
CC   extensive alternative splicing. The alternate splice forms were missing
CC   the imperfect repeats or differed in the number of imperfect repeats. The
CC   human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC   OGFr proteins, nucleic acid molecules, antibodies, transformed cells and
CC   expression vector are useful for detecting expression or levels of an
CC   OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC   cells in vitro. The antisense sequences and antibodies can be used to
CC   promote growth of cells in vitro. Cell growth can be promoted by
CC   interfering with the OGF ligand-receptor system, especially where a
CC   subject suffers from a tissue wound. Treating cancer comprises enhancing
CC   the function of the OGF ligand-receptor system in cancerous cells of a
CC   patient or administering the OGFr nucleic acid to the patient.
XX
SQ   Sequence 1601 BP; 322 A; 485 C; 558 G; 236 T; 0 other;


alignment_scores:
            Quality:   41.00              Length:      9
              Ratio:  4.556                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAA28524/rev    ..

 Align seg 1/1  to reverse of: AAA28524  from: 1  to: 1601

       1 HisGluAlaGluProGluAlaGluPro 9
         |||||||||||||||||||:::::|||
    1418 CACGAAGCAGAGCCTGAGGGGAGCCCG 1392

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28526

seq_documentation_block:
ID   AAA28526 standard; cDNA; 2289 BP.
XX
AC   AAA28526;
XX
DT   29-AUG-2000  (first entry)
```

```
XX
DE   Human opioid growth factor receptor cDNA spliced version 7.
XX
KW   OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW   cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..33
FT                   /*tag=  a
FT   CDS             34..2007
FT                   /*tag=  b
FT                   /product= Opioid_growth_factor_receptor
FT   3'UTR           2008..2289
FT                   /*tag=  c
XX
PN   WO200026340-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-US25802.
XX
PR   03-NOV-1998;   98US-0106879.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Zagon IS,  McLaughlin PJ,  Verderame MF;
XX
DR   WPI; 2000-365594/31.
DR   P-PSDB; AAY92812.
XX
PT   New cDNA encoding rat and human opioid growth factor receptors which
PT   modulate cell growth, useful for treating cancer
XX
PS   Claim 1; Page 87-89; 91pp; English.
XX
CC   Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC   used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC   human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC   consistently yielded a single species of cDNA, while the 3' RACE revealed
CC   extensive alternative splicing. The alternate splice forms were missing
CC   the imperfect repeats or differed in the number of imperfect repeats. The
CC   human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC   OGFr proteins, nucleic acid molecules, antibodies, transformed cells and
CC   expression vector are useful for detecting expression or levels of an
CC   OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC   cells in vitro. The antisense sequences and antibodies can be used to
CC   promote growth of cells in vitro. Cell growth can be promoted by
CC   interfering with the OGF ligand-receptor system, especially where a
CC   subject suffers from a tissue wound. Treating cancer comprises enhancing
CC   the function of the OGF ligand-receptor system in cancerous cells of a
CC   patient or administering the OGFr nucleic acid to the patient.
XX
SQ   Sequence 2289 BP; 470 A; 714 C; 809 G; 296 T; 0 other;

alignment_scores:
           Quality:   41.00            Length:      9
             Ratio:  4.556              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAA28526/rev   ..

 Align seg 1/1  to reverse of: AAA28526  from: 1  to: 2289

       1 HisGluAlaGluProGluAlaGluPro 9
         |||||||||||||||||||:::::|||
    2105 CACGAAGCAGAGCCTGAGGGGAGCCCG 2079

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28522
```

```
seq_documentation_block:
ID   AAA28522 standard; cDNA; 2290 BP.
XX
AC   AAA28522;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human opioid growth factor receptor cDNA of spliced form A.
XX
KW   OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW   cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             34..2007
FT                   /*tag=  a
XX
PN   WO200026340-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-US25802.
XX
PR   03-NOV-1998;   98US-0106879.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Zagon IS,  McLaughlin PJ,  Verderame MF;
XX
DR   WPI; 2000-365594/31.
XX
PT   New cDNA encoding rat and human opioid growth factor receptors which
PT   modulate cell growth, useful for treating cancer
XX
PS   Claim 1; Page 77-78; 91pp; English.
XX
CC   Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC   used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC   human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC   consistently yielded a single species of cDNA, while the 3' RACE revealed
CC   extensive alternative splicing. The alternate splice forms were missing
CC   the imperfect repeats or differed in the number of imperfect repeats. The
CC   human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC   expression vector are useful for detecting expression or levels of an
CC   OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC   cells in vitro. The antisense sequences and antibodies can be used to
CC   promote growth of cells in vitro. Cell growth can be promoted by
CC   interfering with the OGF ligand-receptor system, especially where a
CC   subject suffers from a tissue wound. Treating cancer comprises enhancing
CC   the function of the OGF ligand-receptor system in cancerous cells of a
CC   patient or administering the OGFr nucleic acid to the patient.
XX
SQ   Sequence 2290 BP; 470 A; 713 C; 807 G; 297 T; 3 other;

alignment_scores:
           Quality:   41.00            Length:      9
             Ratio:  4.556              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAA28522/rev   ..

 Align seg 1/1  to reverse of: AAA28522  from: 1  to: 2290

       1 HisGluAlaGluProGluAlaGluPro 9
         |||||||||||||||||||:::::|||
    2106 CACGAAGCAGAGCCTGAGGGGAGCCCG 2080

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28525

seq_documentation_block:
```

```
ID   AAA28525 standard; cDNA; 2348 BP.
XX
AC   AAA28525;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human opioid growth factor receptor cDNA spliced version 4.
XX
KW   OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW   cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..33
FT                   /*tag= a
FT   CDS             34..2067
FT                   /*tag= b
FT                   /product= Opioid_growth_factor_receptor
FT   3'UTR           2068..2348
FT                   /*tag= c
XX
PN   WO200026340-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-US25802.
XX
PR   03-NOV-1998;   98US-0106879.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Zagon IS,  McLaughlin PJ,  Verderame MF;
XX
DR   WPI; 2000-365594/31.
DR   P-PSDB; AAY92811.
XX
PT   New cDNA encoding rat and human opioid growth factor receptors which
PT   modulate cell growth, useful for treating cancer
XX
PS   Claim 1; Page 84-86; 91pp; English.
XX
CC   Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC   used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC   human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC   consistently yielded a single species of cDNA, while the 3' RACE revealed
CC   extensive alternative splicing. The alternate splice forms were missing
CC   the imperfect repeats or differed in the number of imperfect repeats. The
CC   human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC   OGFr proteins, nucleic acid molecules, antibodies, transformed cells and
CC   expression vector are useful for detecting expression or levels of an
CC   OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC   cells in vitro. The antisense sequences and antibodies can be used to
CC   promote growth of cells in vitro. Cell growth can be promoted by
CC   interfering with the OGF ligand-receptor system, especially where a
CC   subject suffers from a tissue wound. Treating cancer comprises enhancing
CC   the function of the OGF ligand-receptor system in cancerous cells of a
CC   patient or administering the OGFr nucleic acid to the patient.
XX
SQ   Sequence 2348 BP; 485 A; 738 C; 826 G; 299 T; 0 other;


alignment_scores:
           Quality:   41.00              Length:      9
             Ratio:  4.556                Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAA28525/rev   ..

 Align seg 1/1  to reverse of: AAA28525  from: 1  to: 2348

       1 HisGluAlaGluProGluAlaGluPro 9
                 |||||||||||||||||:::::|||
    2165 CACGAAGCAGAGCCTGAGGGGAGTCCG 2139

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28523

seq_documentation_block:
ID   AAA28523 standard; cDNA; 2408 BP.
XX
AC   AAA28523;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human opioid growth factor receptor cDNA spliced version 8.
XX
KW   OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW   cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..33
FT                   /*tag= a
FT   CDS             34..2127
FT                   /*tag= b
FT                   /product= Opioid_growth_factor_receptor
FT   3'UTR           2128..2408
FT                   /*tag= c
XX
PN   WO200026340-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-US25802.
XX
PR   03-NOV-1998;   98US-0106879.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Zagon IS,  McLaughlin PJ,  Verderame MF;
XX
DR   WPI; 2000-365594/31.
DR   P-PSDB; AAY92809.
XX
PT   New cDNA encoding rat and human opioid growth factor receptors which
PT   modulate cell growth, useful for treating cancer
XX
PS   Claim 1; Page 79-81; 91pp; English.
XX
CC   Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC   used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC   human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC   consistently yielded a single species of cDNA, while the 3' RACE revealed
CC   extensive alternative splicing. The alternate splice forms were missing
CC   the imperfect repeats or differed in the number of imperfect repeats. The
CC   human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC   expression vector are useful for detecting expression or levels of an
CC   OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC   cells in vitro. The antisense sequences and antibodies can be used to
CC   promote growth of cells in vitro. Cell growth can be promoted by
CC   interfering with the OGF ligand-receptor system, especially where a
CC   subject suffers from a tissue wound. Treating cancer comprises enhancing
CC   the function of the OGF ligand-receptor system in cancerous cells of a
CC   patient or administering the OGFr nucleic acid to the patient.
XX
SQ   Sequence 2408 BP; 500 A; 762 C; 844 G; 302 T; 0 other;


alignment_scores:
           Quality:   41.00              Length:      9
             Ratio:  4.556                Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 77.778

alignment_block:
```

```
US-09-444-281-27 x AAA28523/rev   ..

Align seg 1/1  to reverse of: AAA28523  from: 1  to: 2408

      1 HisGluAlaGluProGluAlaGluPro 9
        ||||||||||||||||||:::::|||
   2225 CACGAAGCAGAGCCTGAGGGGAGCCCG 2199

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH17441

seq_documentation_block:
ID   AAH17441 standard; cDNA; 2409 BP.
XX
AC   AAH17441;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:16891.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 16891; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2409 BP; 495 A; 757 C; 842 G; 315 T; 0 other;

alignment_scores:
            Quality:   41.00              Length:      9
              Ratio:  4.556                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x AAH17441/rev   ..

 Align seg 1/1  to reverse of: AAH17441  from: 1  to: 2409

       1 HisGluAlaGluProGluAlaGluPro 9
         ||||||||||||||||||:::::|||
    2157 CACGAAGCAGAGCCTGAGGGGAGCCCG 2131

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH18239

seq_documentation_block:
ID   AAH18239 standard; cDNA; 5796 BP.
XX
AC   AAH18239;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:18182.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 18182; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
```

CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 5796 BP; 1434 A; 1616 C; 1531 G; 1215 T; 0 other;

alignment_scores:
Quality: 40.00 Length: 8
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-444-281-27 x AAH18239/rev ..

Align seg 1/1 to reverse of: AAH18239 from: 1 to: 5796

1 HisGluAlaGluProGluAlaGlu 8
|||||||||||||||:|||||||
3514 CATGAAGCTGAGCCTCAGGCTGAA 3491

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH18247

seq_documentation_block:
ID AAH18247 standard; cDNA; 6048 BP.
XX
AC AAH18247;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18198.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 18198; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 6048 BP; 1488 A; 1693 C; 1604 G; 1263 T; 0 other;

alignment_scores:
Quality: 40.00 Length: 8
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-444-281-27 x AAH18247/rev ..

Align seg 1/1 to reverse of: AAH18247 from: 1 to: 6048

1 HisGluAlaGluProGluAlaGlu 8
|||||||||||||||:::|||||||
3767 CATGAAGCTGAGCCTCAGGCTGAA 3744

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI61181

seq_documentation_block:
ID AAI61181 standard; cDNA; 8605 BP.
XX
AC AAI61181;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5170.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX

```
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM42025.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 5170; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 8605 BP; 2079 A; 2209 C; 2192 G; 2125 T; 0 other;


alignment_scores:
            Quality:   40.00              Length:      8
              Ratio:  5.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-444-281-27 x AAI61181   ..

 Align seg 1/1  to: AAI61181  from: 1  to: 8605

       1 HisGluAlaGluProGluAlaGlu 8
         ||||||||||||||||:::||||||
    5995 CATGAAGCTGAGCCTCAGGCTGAA 6018

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI59395

seq_documentation_block:
ID   AAI59395 standard; cDNA; 8840 BP.
XX
AC   AAI59395;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1598.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40239.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 1598; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 8840 BP; 2118 A; 2331 C; 2255 G; 2136 T; 0 other;


alignment_scores:
            Quality:   40.00              Length:      8
              Ratio:  5.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-444-281-27 x AAI59395/rev   ..

 Align seg 1/1  to reverse of: AAI59395  from: 1  to: 8840

       1 HisGluAlaGluProGluAlaGlu 8
         ||||||||||||||||:::||||||
    3604 CATGAAGCTGAGCCTCAGGCTGAA 3581

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV68396

seq_documentation_block:
ID   AAV68396 standard; cDNA to mRNA; 9408 BP.
XX
AC   AAV68396;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Human BAZ2-alpha cDNA.
XX
KW   Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW   atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW   BAZ2-beta; treatment; cancer; proliferative disorder; screening; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             740..6376
FT                   /*tag=  a
FT                   /product= "BAZ2-alpha"
```

XX
PN WO9847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP01783.
XX
PR 24-OCT-1997; 97JP-0310027.
PR 18-APR-1997; 97JP-0116570.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;
XX
DR WPI; 1998-583603/49.
DR P-PSDB; AAW81170.
XX
PT Transcriptional regulator gene family containing bromodomain - may
PT be expressed in testis tissue and is useful for treatment of cancer
PT and other proliferative disorders
XX
PS Claim 2; Page 72-88; 187pp; Japanese.
XX
CC This sequence encodes the human BAZ2-alpha protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation).
XX
SQ Sequence 9408 BP; 2279 A; 2458 C; 2383 G; 2280 T; 8 other;

alignment_scores:
Quality: 40.00 Length: 8
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-444-281-27 x AAV68396/rev ..

Align seg 1/1 to reverse of: AAV68396 from: 1 to: 9408

```
       1 HisGluAlaGluProGluAlaGlu 8
         ||||||||||||||:::|||||||
    4343 CATGAAGCTGAGCCTCAGGCTGAA 4320
```

```
OM of: US-09-444-281-27 to: EST:*   out_format : pfs

Date: Jan 4, 2002 10:25 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084143_16183/app_query.fasta_1.210
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09444281_@CGN1_1_3354 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-27
Query length: 11
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 2868.270000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_est2:BF528186       -    46.00   131.87   350.73    619  ! BF528186 602042790F1 NCI_CGAP_B
gb_est1:AW419871       -    45.00   129.34   485.13    593  ! AW419871 fj84f04.y1 zebrafish g
gb_gss:CNS01L9U        -    45.00   126.39   708.38    848  ! AL149363 Anopheles gambiae GSS
gb_gss:AZ208017        -    45.00   125.69   774.85    923  ! AZ208017 SP_0134_B1_D06_T7A Str
gb_est2:BF525048       +    43.00   128.10   569.45    343  ! BF525048 UI-R-AD0-vz-d-06-0-UI.
gb_est1:AW140836       +    43.00   124.17   942.24    552  ! AW140836 EST290918 Normalized r
gb_est2:BG468916       +    43.00   122.72   1.1e+03   658  ! BG468916 602510939F1 NIH_MGC_15
gb_est2:BI295006       -    43.00   122.32   1.2e+03   691  ! BI295006 UI-R-DK0-cee-f-10-0-UI
gb_est2:BF259969       +    43.00   120.74   1.5e+03   836  ! BF259969 HVSMEf0020M02f Hordeum
gb_est1:BE037128       -    43.00   117.44   2.2e+03   1247  ! BE037128 MP15E02 MP Mesembryan
gb_est2:BF655720       +    42.00   130.49   419.04    181  ! BF655720 FM1_48_B05.b1_A003 Flo
gb_est2:T70690         -    42.00   125.71   773.50    323  ! T70690 84 vegetative meristem Z
gb_est2:BI240337       +    42.00   124.09   951.96    393  ! BI240337 RE37162.5prime RE Dros
gb_est1:AU052906       -    42.00   123.56   1.0e+03   419  ! AU052906 AU052906 Dictyostelium
gb_est1:AW923872       +    42.00   123.25   1.1e+03   435  ! AW923872 WS1_30_B09.b1_A002 Wat
gb_est1:AU052382       -    42.00   123.08   1.1e+03   444  ! AU052382 AU052382 Dictyostelium
gb_est1:AU053420       -    42.00   122.93   1.1e+03   452  ! AU053420 AU053420 Dictyostelium
gb_est1:AU039686       -    42.00   122.88   1.1e+03   455  ! AU039686 AU039686 Dictyostelium
gb_est1:AU037503       -    42.00   122.84   1.1e+03   457  ! AU037503 AU037503 Dictyostelium
gb_est1:AU034217       -    42.00   122.44   1.2e+03   480  ! AU034217 AU034217 Dictyostelium
gb_est1:AI512188       +    42.00   122.12   1.2e+03   499  ! AI512188 LD43946.5prime LD Dros
gb_est2:BI239097       +    42.00   121.86   1.3e+03   515  ! BI239097 RE35496.5prime RE Dros
gb_gss:AZ821957        -    42.00   121.79   1.3e+03   519  ! AZ821957 2M0094M21R Mouse 10kb
gb_est1:AU054056       -    42.00   121.73   1.3e+03   523  ! AU054056 AU054056 Dictyostelium
gb_est1:AU052880       -    42.00   121.64   1.3e+03   529  ! AU052880 AU052880 Dictyostelium
gb_est1:AW678908       +    42.00   121.57   1.3e+03   533  ! AW678908 WS1_21_E06.b1_A002 Wat
gb_est1:AI519965       +    42.00   121.53   1.3e+03   536  ! AI519965 LD39848.5prime LD Dros
gb_est2:BI367365       +    42.00   121.51   1.3e+03   537  ! BI367365 RE53119.5prime RE Dros
gb_est1:AI519294       +    42.00   121.48   1.3e+03   539  ! AI519294 LD38826.5prime LD Dros
gb_est1:AU052763       -    42.00   121.23   1.4e+03   556  ! AU052763 AU052763 Dictyostelium
gb_est1:AU053309       -    42.00   121.14   1.4e+03   562  ! AU053309 AU053309 Dictyostelium
gb_est2:BI215736       +    42.00   121.08   1.4e+03   566  ! BI215736 RE22333.5prime RE Dros
gb_est1:AI405045       +    42.00   121.02   1.4e+03   570  ! AI405045 GH24953.5prime GH Dros
gb_est1:AA990672       +    42.00   120.98   1.4e+03   573  ! AA990672 LD34484.5prime LD Dros
gb_est2:BI234021       +    42.00   120.91   1.4e+03   578  ! BI234021 RE29638.5prime RE Dros
gb_est1:AI517604       +    42.00   120.89   1.4e+03   579  ! AI517604 GH28630.5prime GH Dros
gb_est1:AI387826       +    42.00   120.86   1.4e+03   581  ! AI387826 GH18490.5prime GH Dros
gb_est2:BI227816       +    42.00   120.83   1.4e+03   583  ! BI227816 RE25083.5prime RE Dros
gb_est2:BI365222       +    42.00   120.81   1.5e+03   585  ! BI365222 RE50464.5prime RE Dros
gb_est1:AI518766       +    42.00   120.76   1.5e+03   588  ! AI518766 LD38476.5prime LD Dros
gb_est1:AI531629       +    42.00   120.75   1.5e+03   589  ! AI531629 SD02789.5prime SD Dros
gb_est1:BE345652       -    42.00   120.72   1.5e+03   591  ! BE345652 946021H07.y2 946 -
gb_est2:BI370648       +    42.00   120.67   1.5e+03   595  ! BI370648 RE57356.5prime RE D
gb_est1:BE345083       -    42.00   120.56   1.5e+03   603  ! BE345083 946031D04.y1 946 -
gb_est1:AI518800       +    42.00   120.54   1.5e+03   604  ! AI518800 LD38583.5prime LD D

seq_name: gb_est2:BF528186

seq_documentation_block:
LOCUS       BF528186      619 bp    mRNA            EST       11-DEC-2000
DEFINITION  602042790F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4180198
            5', mRNA sequence.
ACCESSION   BF528186.
VERSION     BF528186.1  GI:11615549
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9491  row: d  column: 23
            High quality sequence stop: 350.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4180198"
                     /clone_lib="NCI_CGAP_Brn67"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      115 a    219 c    213 g     72 t
ORIGIN


alignment_scores:
            Quality:  46.00              Length:      9
              Ratio:  5.111                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-27 x BF528186/rev   ..

 Align seg 1/1  to reverse of: BF528186  from: 1  to: 619

       1 HisGluAlaGluProGluAlaGluPro 9
         |||||||||||||||||:::|||||||
     525 CACGAAGCAGAGCCTGAAGGGGAGCCC 499

seq_name: gb_est1:AW419871

seq_documentation_block:
LOCUS       AW419871      593 bp    mRNA            EST       09-FEB-2000
DEFINITION  fj84f04.y1 zebrafish gridded kidney Danio rerio cDNA 5' similar to
            TR:Q13436 Q13436 NUCLEOLAR RNA HELICASE GU ;, mRNA sequence.
ACCESSION   AW419871
VERSION     AW419871.1  GI:6947803
KEYWORDS    EST.
```

```
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fj84f04.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, Missouri (web address:
            www.genomesystems.com) (email contact: info@genomesystems.com) and
            Research Genetics, Huntsville, Alabama (web address: www.resgen.com
            ) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="zebrafish gridded kidney"
                     /sex="mixed"
                     /tissue_type="kidney pooled from 300 wild type adults"
                     /lab_host="XLOLR"
                     /note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI;
                     Site_2: XhoI; Oligo dT cDNA library constructed from mRNA
                     pooled from pooled kidney  tissue from 300 adult
                     zebrafish."
BASE COUNT      138 a    153 c    181 g    121 t
ORIGIN


alignment_scores:
            Quality:   45.00            Length:     11
              Ratio:  4.500              Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x AW419871/rev  ..

 Align seg 1/1  to reverse of: AW419871  from: 1  to: 593

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         ||||||:::|||||||||   ||||||::::::
     276 CACGAGTCGGAGCCAGAACCAGAACCTTTGGTG 244

seq_name: gb_gss:CNS01L9U

seq_documentation_block:
LOCUS       CNS01L9U      848 bp    DNA             GSS       14-JUN-2001
DEFINITION  Anopheles gambiae GSS T7 end of clone 18M20 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL149363
VERSION     AL149363.1  GI:7009842
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Anopheles.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 848)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .848
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="18M20"
                     /clone_lib="NotreDame1"
                     /note="end : T7"
BASE COUNT      216 a    157 c    204 g    248 t     23 others
ORIGIN


alignment_scores:
            Quality:   45.00            Length:     11
              Ratio:  4.500              Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 72.727

alignment_block:
 US-09-444-281-27 x CNS01L9U/rev  ..

 Align seg 1/1  to reverse of: CNS01L9U  from: 1  to: 848

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         ||||||:::   |||||||:::|||||||||||
     666 CATGAATCCATACCTGAATCGGAACCTATCATG 634

seq_name: gb_gss:AZ208017

seq_documentation_block:
LOCUS       AZ208017      923 bp    DNA             GSS       31-AUG-2000
DEFINITION  SP_0134_B1_D06_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=134 Col=11 Row=H, DNA sequence.
ACCESSION   AZ208017
VERSION     AZ208017.1  GI:8420201
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
```

```
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 134  row: H  column: 11
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 923.
FEATURES              Location/Qualifiers
     source           1. .923
                      /organism="Strongylocentrotus purpuratus"
                      /db_xref="taxon:7668"
                      /clone="Plate=134 Col=11 Row=H"
                      /clone_lib="Strongylocentrotus purpuratus, purple sea
                      urchin, sperm genomic BAC library"
                      /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                      DH10B"
BASE COUNT      173 a    216 c    203 g    331 t
ORIGIN


alignment_scores:
            Quality:   45.00          Length:      9
              Ratio:  5.625            Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-444-281-27 x AZ208017/rev   ..

 Align seg 1/1  to reverse of: AZ208017  from: 1  to: 923

       1 HisGluAlaGluProGluAlaGluPro 9
         ||||||||||||||||||   ||||||
      55 CACGAAGCCGAGCCAGAGCCAGAACCA 29

seq_name: gb_est2:BF525048

seq_documentation_block:
LOCUS        BF525048      343 bp    mRNA            EST       11-DEC-2000
DEFINITION   UI-R-AD0-vz-d-06-0-UI.r1 UI-R-AD0 Rattus norvegicus cDNA clone
             UI-R-AD0-vz-d-06-0-UI 5', mRNA sequence.
ACCESSION    BF525048
VERSION      BF525048.1  GI:11633015
KEYWORDS     EST.
SOURCE       Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 343)
  AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE      Normalization and subtraction: two approaches to facilitate gene
             discovery
  JOURNAL    Genome Res. 6 (9), 791-806 (1996)
  MEDLINE    97044477
COMMENT      Contact: Soares, MB
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: msoares@blue.weeg.uiowa.edu
             cDNA Library Preparation: M.B. Soares Lab Clone distribution:
             clones will be available through Research Genetics (www.resgen.com)
             This clone is also available through the I.M.A.G.E. Consortium at
             LLNL (info@image.llnl.gov). IMAGE ID= 1794853
             Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .343
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-AD0-vz-d-06-0-UI"
                      /clone_lib="UI-R-AD0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AD0
                      library is a non-normalized library constructed from  15
                      dpc rat atrium.  The tag is a string of 5 nucleotides
                      present  between the Not I site and the oligo-dT track.
                      The library was  constructed as described by Bonaldo,
                      Lennon and Soares,  Genome Research 6: 791-806, 1996.
                      Tissue provided by Jim Lin, Department of Biology,
                      University of Iowa."
BASE COUNT      106 a     83 c     95 g     59 t
ORIGIN


alignment_scores:
            Quality:   43.00          Length:     11
              Ratio:  3.909            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x BF525048   ..

 Align seg 1/1  to: BF525048  from: 1  to: 343

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||:::||||||||||:::|||||||||::::::
     127 CATCATGCAGAGCCTCATGCAGAACCTCTTGTC 159

seq_name: gb_est1:AW140836

seq_documentation_block:
LOCUS        AW140836      552 bp    mRNA            EST       30-OCT-1999
DEFINITION   EST290918 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
             RGIAV22 5' end similar to peroxisomal farnesylated protein, mRNA
             sequence.
ACCESSION    AW140836
VERSION      AW140836.1  GI:6160674
KEYWORDS     EST.
SOURCE       Rattus sp.
  ORGANISM   Rattus sp.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 552)
  AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
             Kerlavage,A.R. and Adams,M.D.
  TITLE      Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
             Gene Index
  JOURNAL    Unpublished (1998)
COMMENT      Other_ESTs: TC85992
             Contact: Lee, NH
             The Institute for Genomic Research
             9712, Medical Center Drive, Rockville, MD 20850, USA
             Tel: (301)-838-3529
             Fax: (301)-838-0208
             Email: nhlee@tigr.org
             For clone availability, additional sequence and expression
             information related to this EST please check the TIGR Rat Gene
             Index (http://www.tigr.org/tdb/rgi/rgi.html). To order a clone
             contact the ATCC (http://www.atcc.org/atcc.html).
             Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .552
                      /organism="Rattus sp."
                      /db_xref="taxon:10118"
                      /clone="RGIAV22"
                      /clone_lib="Normalized rat heart, Bento Soares"
                      /note="Organ: heart; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
BASE COUNT      155 a    138 c    149 g    110 t
ORIGIN
```

```
alignment_scores:
            Quality:   43.00              Length:     11
              Ratio:  3.909                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x AW140836   ..

 Align seg 1/1  to: AW140836  from: 1  to: 552

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||:::|||||||||:::|||||||||:::::
     483 CATCATGCAGAGCCTCATGCAGAACCTCTTGTC 515

seq_name: gb_est2:BG468916

seq_documentation_block:
LOCUS       BG468916      658 bp    mRNA            EST       21-MAR-2001
DEFINITION  602510939F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4645424 5',
            mRNA sequence.
ACCESSION   BG468916
VERSION     BG468916.1  GI:13401293
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1418  row: m  column: 09
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .658
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4645424"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      210 a    152 c    172 g    124 t
ORIGIN


alignment_scores:
            Quality:   43.00              Length:     10
              Ratio:  5.375                Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x BG468916   ..

 Align seg 1/1  to: BG468916  from: 1  to: 658

       1 HisGluAlaGluProGluAlaGluProIle 10
         |||||||   |||||||||   |||||||:::
     390 CACGAACCTGAGCCAGAACCTGAGCCTGTT 419

seq_name: gb_est2:BI295006

seq_documentation_block:
LOCUS       BI295006      691 bp    mRNA            EST       19-JUL-2001
DEFINITION  UI-R-DK0-cee-f-10-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
            UI-R-DK0-cee-f-10-0-UI 3', mRNA sequence.
ACCESSION   BI295006
VERSION     BI295006.1  GI:14958018
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized rat heart pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-DK0-cee-f-10-0-UI"
                     /clone_lib="UI-R-DK0"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
                     library is a subtracted library derived from a mixture of
                     five individually tagged normalized rat libraries:
                     brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
                     aorta-nRAP (20%), and placenta-nRPP (20%). Each original
                     library was constructed from a mixture of equal amounts of
                     RNA from seven different developmental time-points:
                     embryonic day 17, embryonic day 19, embryonic day 21,
                     adult day 1, adult day 12, adult day 75, and adult day
                     200. (Exception: the aorta pool does not contain embryonic
                     day 17 RNA and the placenta pool contains only the three
                     embryonic stages). Each library was normalized
                     individually according to the procedure described by
                     Bonaldo, Lennon & Soares (Genome Research Genome 6:
                     791-806, 1996). For construction of the DK0 subtracted
                     library, plasmid DNA from each of the five individually
                     tagged normalized libraries was mixed in the proportions
                     specified above and electroporated into competent bacteria
                     for production of single-stranded circular DNA
                     representing the pool of libraries. Single-stranded
                     circular DNA representing these five normalized libraries
                     was then used as a tracer in a subtractive hybridization
                     with a driver (PCR amplified inserts from a plasmid DNA
```

```
                     template preparation) comprising: a) a set of about 1,000
                     arrayed clones from each of the five non-normalized
                     libraries of brain (CT0s), heart (CS0s), kidney (CU0s),
                     aorta (CW0s), and placenta (CX0s). The resulting pool of
                     approximately 5,000 clones represented about 33.3% of the
                     final driver population. A set of about 2,000 arrayed
                     clones from each of the five normalized libraries of brain
                     (CT0), heart (CS0), kidney (CU0), aorta (CW0), and
                     placenta (CX0). The resulting pool of about 10,000 clones
                     represented about 66.6% of the final driver population.
                     TAG_LIB=UI-R-DK0
                     TAG_TISSUE=rat heart pool
                     TAG_SEQ=ATAAGATAAC"
BASE COUNT      126 a    179 c    169 g    217 t
ORIGIN


alignment_scores:
            Quality:   43.00             Length:     11
              Ratio:  3.909               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x BI295006/rev   ..

 Align seg 1/1  to reverse of: BI295006  from: 1  to: 691

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||:::||||||||||:::|||||||||::::::
     435 CATCATGCAGAGCCTCATGCAGAACCTCTTGTC 403

seq_name: gb_est2:BF259969

seq_documentation_block:
LOCUS       BF259969      836 bp    mRNA            EST       23-FEB-2001
DEFINITION  HVSMEf0020M02f Hordeum vulgare seedling root EST library HVcDNA0007
            (etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMEf0020M02f, mRNA sequence.
ACCESSION   BF259969
VERSION     BF259969.2  GI:13120375
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     On Nov 16, 2000 this sequence version replaced gi:11189082.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEf0020M02f"
                     /clone_lib="Hordeum vulgare seedling root EST library
                     HVcDNA0007 (etiolated and unstressed)"
                     /tissue_type="seedling root"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; For
                     more details on library preparation and sequence analysis
                     see http://www.genome.clemson.edu/projects/barley/ To
                     order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      163 a    233 c    275 g    165 t
ORIGIN


alignment_scores:
            Quality:   43.00             Length:      9
              Ratio:  4.778               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-27 x BF259969   ..

 Align seg 1/1  to: BF259969  from: 1  to: 836

       1 HisGluAlaGluProGluAlaGluPro 9
         |||:::|||||||||:::|||||||||
     614 CATGATGCTGAACCGAAAGCTGAGCCG 640

seq_name: gb_est1:BE037128

seq_documentation_block:
LOCUS       BE037128     1247 bp    mRNA            EST       07-JUN-2000
DEFINITION  MP15E02 MP Mesembryanthemum crystallinum cDNA 5' similar to
            auxin-regulated protein, mRNA sequence.
ACCESSION   BE037128
VERSION     BE037128.1  GI:8332144
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 1247)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu.
FEATURES             Location/Qualifiers
     source          1. .1247
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MP"
                     /tissue_type="apical meristem and leaf primordia"
                     /dev_stage="6 weeks"
                     /note="3 d 500mM NaCl"
BASE COUNT      270 a    300 c    242 g    301 t    134 others
ORIGIN


alignment_scores:
            Quality:   43.00             Length:     10
              Ratio:  5.375               Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x BE037128/rev   ..

 Align seg 1/1  to reverse of: BE037128  from: 1  to: 1247

       1 HisGluAlaGluProGluAlaGluProIle 10
         ||||||   |||||||||   |||||||:::
```

```
     639 CATGAACCCGAACCCGAACCCGAACCCGTC 610

seq_name: gb_est2:BF655720

seq_documentation_block:
LOCUS       BF655720      181 bp    mRNA            EST       20-DEC-2000
DEFINITION  FM1_48_B05.b1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
            propinquum cDNA, mRNA sequence.
ACCESSION   BF655720
VERSION     BF655720.1  GI:11920852
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 181)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: floral-induced meristems
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 111
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .181
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Floral-Induced Meristem 1 (FM1)"
                     /note="Organ: Floral-induced meristems; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; mature plants were placed in a growth chamber for
                     15 days with 16 hr darkness and 8 hr light (flowering is
                     induced by short-day conditions); 16 days after being
                     returned to the greenhouse under natural long days during
                     late April/early May, meristems were harvested The
                     library was made from poly-A RNA in the cloning vector
                     lambda ZAP II. Clones to be sequenced were prepared by
                     mass excision."
BASE COUNT       52 a     40 c     61 g     28 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:     11
              Ratio:  4.200              Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x BF655720   ..

 Align seg 1/1  to: BF655720  from: 1  to: 181

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||:::|||:::|||||||||||||   |||:::
     107 CATCAGGCGCAACCGGAAGCGGAAGCCATCCTA 139

seq_name: gb_est2:T70690

seq_documentation_block:
LOCUS       T70690        323 bp    mRNA            EST       17-OCT-1996
DEFINITION  84 vegetative meristem Zea mays cDNA clone 7C03B11, mRNA sequence.
ACCESSION   T70690
VERSION     T70690.1  GI:681838
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Shen,B., Carneiro,N., Torres-Jerez,I., Stevenson,R., McCreery,T.,
            Helentjaris,T., Baysdorfer,C., Almira,E., Ferl,R., Habben,J. and
            Larkins,B.
  TITLE     Partial sequencing and mapping of clones from two maize cDNA
            libraries
  JOURNAL   Plant Mol. Biol. 26, 1085-1101 (1994)
  MEDLINE   95111093
COMMENT     Contact: The Maize cDNA Project

            Helentjaris TG (primary contact)
            Pioneer Hi-Bred Int., Inc.
            Agronomic Traits/TTD
            7300 N.W. 62nd Ave. P.O. Box 1004
            Johnston, IA 50131-1004
            ph: 515-270-3691
            fax: 515-270-4312
            E-mail: helnjartg@phibred.com

            Chris Baysdorfer
            Department of Biological Sciences, School of Science
            California State University, Hayward
            Hayward, CA 94542
            ph: 510-881-3459
            fax: 510-727-2035
            E-mail: cbaysdor@s1.csuhayward.edu

            Rob Ferl
            Interdisciplinary Center for Biotechnology Research
            DNA Sequencing Core
            University of Florida
            P.O. Box 100695
            Gainesville, FL 32611-0695
            ph: 904-392-1928, ext. 301
            fax: 904-392-4072
            E-mail: robferlnervm.nerdc.ufl.edu

           Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Zea mays"
                     /strain="W64A2"
                     /db_xref="taxon:4577"
                     /clone="7C03B11"
                     /clone_lib="vegetative meristem"
                     /lab_host="SOLR"
                     /note="Vector: Zap; Site_1: EcoRI; Site_2: XhoI; ds-cDNA
                     was prepared from oligo-dT selected mRNA by priming with a
                     XhoI oligo- dT oligomer and then adding the second strand
                     to RNase-nicked DNA:RNA hybrid with DNA PolI.  EcoRI
                     adaptors were added to the ends, the ds-cDNAs were then
                     digested with EcoRI- XhoI and size-selected. These were
                     directionally-cloned into the ZAP phage vector, excised as
                     plasmids, and then individually analyzed."
BASE COUNT       55 a     91 c     97 g     64 t     16 others
ORIGIN


alignment_scores:
            Quality:   42.00            Length:     10
              Ratio:  5.250              Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x T70690/rev   ..
```

```
 Align seg 1/1  to revers  of: T70690  from: 1  to: 323

       1 HisGluAlaGluProGluAlaGluProIle 10
         ||||||   |||||||||   ||||||:::
      36 CACGAACCAGAACCTGAACACGAACCGGTG 7

seq_name: gb_est2:BI240337

seq_documentation_block:
LOCUS       BI240337      393 bp    mRNA            EST       12-JUL-2001
DEFINITION  RE37162.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE37162 5 similar to CG7611:
            FBan0007611 located on: 3L 78E1-78E1; : 05/12/2001, mRNA sequence.
ACCESSION   BI240337
VERSION     BI240337.1  GI:14708940
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003594: arm:3L [21112610,21418125]
            estimated-cyto:78C7-78E3: 05/12/2001
            Plate: RE.371  row: F  column: 2
            High quality sequence stop: 342.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE37162"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT       99 a    106 c    101 g     87 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:      8
              Ratio:  5.250              Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
 US-09-444-281-27 x BI240337

 Align seg 1/1  to: BI240337  from: 1  to: 393

       2 GluAlaGluProGluAlaGluPro 9
         |||||||||||||||||||||||||
     190 GAGGCTGAGCCTGAGGCAGAGCCT 213

seq_name: gb_est1:AU052906

seq_documentation_block:
LOCUS       AU052906      419 bp    mRNA            EST       28-APR-1999
DEFINITION  AU052906 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLF303, mRNA sequence.
ACCESSION   AU052906
VERSION     AU052906.1  GI:4701389
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLF303"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      118 a     86 c     89 g    126 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:      8
              Ratio:  5.250              Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
 US-09-444-281-27 x AU052906/rev

 Align seg 1/1  to reverse of: AU052906  from: 1  to: 419

       2 GluAlaGluProGluAlaGluPro 9
         |||||||||||||||||||||||||
     164 GAGGCGGAACCTGAGGCGGAACCT 141

seq_name: gb_est1:AW923872

seq_documentation_block:
LOCUS       AW923872      435 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_30_B09.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW923872
VERSION     AW923872.1  GI:8089697
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
```

Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 368
POLYA=No.

```
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      156 a    162 c     84 g     33 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:     10
              Ratio:  4.667              Gaps:        0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x AW923872

 Align seg 1/1  to: AW923872  from: 1  to: 435

       1 HisGluAlaGluProGluAlaGluProIle 10
         ||||||||||||||:::   :::||||||
       2 CACGAGGCTGAGCCTCAACCTAAACCCATA 31
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:40:32 ; Search time 50.17 Seconds
(without alignments)
32.071 Million cell updates/sec

Title: US-09-444-281-27
Perfect score: 59
Sequence: 1 HEAEPEAEPIM 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	71.2	880	5	O17338	O17338 caenorhabdi
2	42	71.2	885	5	Q9TXR9	Q9txr9 caenorhabdi
3	42	71.2	930	5	O17339	O17339 caenorhabdi
4	41	69.5	874	5	Q9VQL0	Q9vql0 drosophila
5	40	67.8	248	12	Q65200	Q65200 african swi
6	40	67.8	1380	5	Q9VIU8	Q9viu8 drosophila
7	39	66.1	251	3	Q9USZ4	Q9usz4 schizosacch
8	39	66.1	304	5	Q9NAB8	Q9nab8 caenorhabdi
9	39	66.1	340	5	Q9NAA2	Q9naa2 caenorhabdi
10	39	66.1	408	5	O96979	O96979 paranophrys
11	39	66.1	417	5	Q9W4Y3	Q9w4y3 drosophila
12	39	66.1	435	5	Q9NEF8	Q9nef8 drosophila
13	39	66.1	848	5	Q21489	Q21489 caenorhabdi
14	39	66.1	1192	5	Q9W475	Q9w475 drosophila
15	39	66.1	1277	12	Q08547	Q08547 cercopithec
16	39	66.1	1279	12	Q66031	Q66031 cercopithec
17	39	66.1	1458	3	Q9HE72	Q9he72 neurospora
18	39	66.1	1593	5	Q20207	Q20207 caenorhabdi
19	39	66.1	6797	2	Q9X993	Q9x993 streptomyce
20	38	64.4	150	1	Q9YFA9	Q9yfa9 aeropyrum p
21	38	64.4	200	4	Q9UQD8	Q9uqd8 homo sapien
22	38	64.4	288	3	Q9P3H2	Q9p3h2 neurospora
23	38	64.4	592	3	O59900	O59900 cryptococcu
24	38	64.4	682	2	Q9A8Q8	Q9a8q8 caulobacter
25	38	64.4	1083	2	O86637	O86637 streptomyce
26	37	62.7	20	4	Q9UD25	Q9ud25 homo sapien
27	37	62.7	101	6	Q95284	Q95284 sus scrofa
28	37	62.7	125	2	Q9WX24	Q9wx24 streptomyce
29	37	62.7	175	10	Q9SVE4	Q9sve4 arabidopsis
30	37	62.7	208	4	O60937	O60937 homo sapien
31	37	62.7	248	11	Q63224	Q63224 rattus norv
32	37	62.7	251	6	Q9GLN2	Q9gln2 bos taurus
33	37	62.7	265	2	Q9JZP0	Q9jzp0 neisseria m
34	37	62.7	265	2	Q9JUS3	Q9jus3 neisseria m
35	37	62.7	338	10	Q9M4K6	Q9m4k6 anthoceros
36	37	62.7	338	10	Q9M4K5	Q9m4k5 anthoceros
37	37	62.7	356	10	O64647	O64647 arabidopsis
38	37	62.7	387	5	Q9VZC7	Q9vzc7 drosophila
39	37	62.7	401	5	Q9VZC6	Q9vzc6 drosophila
40	37	62.7	407	5	Q9NGB6	Q9ngb6 drosophila
41	37	62.7	439	11	Q9JMF5	Q9jmf5 mus musculu
42	37	62.7	439	11	Q9DC79	Q9dc79 mus musculu
43	37	62.7	583	4	Q9UKV2	Q9ukv2 homo sapien
44	37	62.7	612	5	Q9U003	Q9u003 caenorhabdi
45	37	62.7	614	4	Q9UKV1	Q9ukv1 homo sapien

ALIGNMENTS

```
RESULT 1
O17338
ID   O17338      PRELIMINARY;      PRT;   880 AA.
AC   O17338;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   T23E7.2B PROTEIN.
GN   T23E7.2B.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson.A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Latreille P., Stellyes L., Elliott G.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026205; AAB71258.1; -.
SQ   SEQUENCE   880 AA;  95398 MW;  97A8A101E8FBA1C1 CRC64;
```

Query Match
Best Local Similarity 71.2%; Score 42; DB 5; Length 880;
Matches 8; Conservative 100.0%; Pred. No. 33;
Qy 2 EAEPEAEP 9 Mismatches 0; Indels 0; Gaps 0;
Db 469 EAEPEAEP 476

RESULT 2
Q9TXR9
ID Q9TXR9
AC Q9TXR9; PRELIMINARY; PRT; 885 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE T23E7.2C PROTEIN.
GN T23E7.2C.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OX Rhabditidae; Peloderinae; Caenorhabditis.
RN NCBI_TaxID=6239;
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RA MEDLINE=99069613; PubMed=9851916;
RT None;
RT "Genome sequence of the nematode C. elegans: a platform for
RL investigating biology. The C. elegans Sequencing Consortium.";
RN Science 282:2012-2018(1998).
RP [2]
RC SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RT Latreil P., Stellyes L., Elliot, G, Wilson R.;
RL "The sequence of C. elegans cosmid T23E7.";
RN Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RP [3]
RC SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RL Waterston R.;
DR Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF026205; AAD47129.1; -.
DR InterPro; IPR002965; Prich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 885 AA; 96409 MW; 0ABAF598C91507D9 CRC64;

Query Match
Best Local Similarity 71.2%; Score 42; DB 5; Length 885;
Matches 8; Conservative 100.0%; Pred. No. 33;
Qy 2 EAEPEAEP 9 Mismatches 0; Indels 0; Gaps 0;
Db 519 EAEPEAEP 526

RESULT 3
O17339
ID O17339
AC O17339; PRELIMINARY; PRT; 930 AA.
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE T23E7.2B PROTEIN.
GN T23E7.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OX Rhabditidae; Peloderinae; Caenorhabditis.
RN NCBI_TaxID=6239;
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A ,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L ,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R ,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RT Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RN Nature 368:32-38(1994).
RP [2]
RC SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RL Latreille P., Stellyes L., Elliott G.;
RN Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RP [3]
RC SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RL Waterston R.;
DR Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
SQ EMBL; AF026205; AAB71257.1; -.
SEQUENCE 930 AA; 101210 MW; 50B188B630B34ABD CRC64;

Query Match
Best Local Similarity 71.2%; Score 42; DB 5; Length 930;
Matches 8; Conservative 100.0%; Pred. No. 35;
Qy 2 EAEPEAEP 9 Mismatches 0; Indels 0; Gaps 0;
Db 519 EAEPEAEP 526

RESULT 4
Q9VQL0
ID Q9VQL0
AC Q9VQL0; PRELIMINARY; PRT; 874 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG17258 PROTEIN.
GN CG17258.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
RN NCBI_TaxID=7227;
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W ,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W M , Glasser K ,
RA Glodek A., Gong F., Gorrell J.H., Gu Z , Guan P , Harris M ,

```
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003581; AAF51156.1; -.
DR   FlyBase; FBgn0031496; CG17258.
SQ   SEQUENCE   874 AA;  103694 MW;  5F56D1C0E7A01D9A CRC64;


  Query Match             69.5%;  Score 41;  DB 5;  Length 874;
  Best Local Similarity   70.0%;  Pred. No. 49;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPIM 11
           ||||| ||::
Db     376 EAEPEQEPLL 385


RESULT   5
Q65200
ID   Q65200      PRELIMINARY;      PRT;   248 AA.
AC   Q65200;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PE248R.
GN   E248R.
OS   African swine fever virus (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
OC   African swine fever-like viruses.
OX   NCBI_TaxID=10497;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=96036500; PubMed=7483270;
RA   Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA   Rodriguez J.F., Vinuela E.;
RT   "Immune protection conferred by the baculovirus-related glycoprotein
RT   of Thogoto virus (Orthomyxoviridae).";
RL   Virology 208:249-278(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=94233765; PubMed=8178480;
RA   La Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;
RT   "Nucleotide sequence and variability of the inverted terminal
RT   repetitions of African swine fever virus DNA.";
RL   Virology 201:152-156(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=90219205; PubMed=2325203;
RA   Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
RA   La Vega I., Blasco R., Vinuela E.;
RT   "Multigene families in African swine fever virus: family 360.";
RL   J. Virol. 64:2073-2081(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=90219204; PubMed=2325202;
RA   Almendral J.M., Almazan F., Blasco R., Vinuela E.;
RT   "Multigene families in African swine fever virus: family 110.";
RL   J. Virol. 64:2064-2072(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=91134988; PubMed=1994575;
RA   Camacho A., Vinuela E.;
RT   "Protein p22 of African swine fever virus: an early structural protein
RT   that is incorporated into the membrane of infected cells.";
RL   Virology 181:251-257(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RA   Almazan F., Murguia J.R., Rodriguez J.M., La Vega I., Vinuela E.;
RL   J. Gen. Virol. 0:0-0(0).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=94187118; PubMed=8139051;
RA   Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,
RA   Vinuela E.;
RT   "Multigene families in African swine fever virus: family 505.";
RL   J. Virol. 68:2746-2751(1994).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93346971; PubMed=8393914;
RA   Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
RT   "African swine fever virus thymidylate kinase gene: sequence and
RT   transcriptional mapping.";
RL   J. Gen. Virol. 74:1633-1638(1993).
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=94065656; PubMed=8245848;
RA   Alcami A., Angulo A., Vinuela E.;
RT   "Mapping and sequence of the gene encoding the African swine fever
RT   virion protein of M(r) 11500.";
RL   J. Gen. Virol. 74:2317-2324(1993).
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93277388; PubMed=8503790;
RA   Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
RT   "Structure and expression in E. coli of the gene coding for protein
RT   p10 of African swine fever virus.";
RL   Arch. Virol. 130:93-107(1993).
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=90357780; PubMed=2389555;
RA   Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,
RA   Vinuela E.;
RT   "Sequence and evolutionary relationships of African swine fever virus
RT   thymidine kinase.";
RL   Virology 178:301-304(1990).
RN   [12]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93281390; PubMed=8506138;
RA   Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
RT   "African swine fever virus encodes two genes which share significant
RT   homology with the two largest subunits of DNA-dependent RNA
RT   polymerases.";
RL   Nucleic Acids Res. 21:2423-2427(1993).
RN   [13]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
```

```
RX   MEDLINE=93353606; PubMed=8102411;
RA   Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
RT   "African swine fever virus encodes a CD2 homolog responsible for the
RT   adhesion of erythrocytes to infected cells.";
RL   J. Virol. 67:5312-5320(1993).
RN   [14]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=94085774; PubMed=8262374;
RA   Yanez R.J., Rodriguez J.M., Boursnell M., Rodriguez J.F., Vinuela E.;
RT   "Two putative African swine fever virus helicases similar to yeast
RT   'DEAH' pre-mRNA processing proteins and vaccinia virus ATPases D11L
RT   and D6R.";
RL   Gene 134:161-174(1993).
RN   [15]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=90223993; PubMed=2327074;
RA   Lopez-Otin C., Freije J.M., Parra F., Mendez E., Vinuela E.;
RT   "Mapping and sequence of the gene coding for protein p72, the major
RT   capsid protein of African swine fever virus.";
RL   Virology 175:477-484(1990).
RN   [16]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=94123986; PubMed=8293992;
RA   Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
RT   "The DNA polymerase-encoding gene of African swine fever virus:
RT   sequence and transcriptional analysis.";
RL   Gene 136:103-110(1993).
RN   [17]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93327788; PubMed=8335009;
RA   Simon-Mateo C., Andres G., Vinuela E.;
RT   "Polyprotein processing in African swine fever virus: a novel gene
RT   expression strategy for a DNA virus.";
RL   EMBO J. 12:2977-2987(1993).
RN   [18]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93233210; PubMed=8474154;
RA   Prados F.J., Vinuela E., Alcami A.;
RT   "Sequence and characterization of the major early phosphoprotein p32
RT   of African swine fever virus.";
RL   J. Virol. 67:2475-2485(1993).
RN   [19]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=92260660; PubMed=1583732;
RA   Alcami A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.,
RA   Carrascosa A.L., Vinuela E.;
RT   "Amino acid sequence and structural properties of protein p12, an
RT   African swine fever virus attachment protein.";
RL   J. Virol. 66:3860-3868(1992).
RN   [20]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93174976; PubMed=8438592;
RA   Yanez R.J., Vinuela E.;
RT   "African swine fever virus encodes a DNA ligase.";
RL   Virology 193:531-536(1993).
RN   [21]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=93174941; PubMed=8382399;
RA   Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
RT   "African swine fever virus guanylyltransferase.";
RL   Virology 193:319-328(1993).
RN   [22]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=95159428; PubMed=7856088;
RA   Simon-Mateo C., Freije J.M., Andres G., Lopez-Otin C., Vinuela E.;
RT   "Mapping and sequence of the gene encoding protein p17, a major
RT   African swine fever virus structural protein.";
RL   Virology 206:1140-1144(1995).
RN   [23]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=92263807; PubMed=1316688;
RA   Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.,
RA   Salas M.L.;
RT   "A gene homologous to topoisomerase II in African swine fever virus.";
RL   Virology 188:938-947(1992).
RN   [24]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BA71V;
RX   MEDLINE=94091056; PubMed=8266720;
RA   Freije J.M., Lain S., Vinuela E., Lopez-Otin C.;
RT   "Nucleotide sequence of a nucleoside triphosphate phosphohydrolase

  Query Match             67.8%;  Score 40;  DB 12;  Length 248;
  Best Local Similarity   63.6%;  Pred. No. 20;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           || | ||||::
Db     227 HEEEEEAEPLI 237


RESULT   6
Q9VIU8
ID   Q9VIU8      PRELIMINARY;      PRT;  1380 AA.
AC   Q9VIU8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG10132 PROTEIN.
GN   CG10132.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

```
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003663; AAF53816.1; -.
DR   FlyBase; FBgn0032798; CG10132.
DR   InterPro; IPR001680; WD40.
DR   SMART; SM00320; WD40; 2.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   1380 AA;  154423 MW;  C1928D066450A15B CRC64;


  Query Match             67.8%;  Score 40;  DB 5;  Length 1380;
  Best Local Similarity   66.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEP 9
           |::||| ||
Db     179 HDSEPETEP 187


RESULT   7
Q9USZ4
ID   Q9USZ4      PRELIMINARY;      PRT;   251 AA.
AC   Q9USZ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 28.0 KDA PROTEIN.
GN   SPBC11G11.05.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Saunders D., Harris D., McDougall R.C., Rajandream M.A., Barrell B.G.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132717; CAB59807.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   251 AA;  28010 MW;  8018E8325AC65B99 CRC64;


  Query Match             66.1%;  Score 39;  DB 3;  Length 251;
  Best Local Similarity   87.5%;  Pred. No. 30;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
           |:||||||
Db     185 ESEPEAEP 192


RESULT   8
Q9NAB8
ID   Q9NAB8      PRELIMINARY;      PRT;   304 AA.
AC   Q9NAB8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Y53F4B.16 PROTEIN.
GN   Y53F4B.16.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smye R.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL132949; CAB61086.1; -.
SQ   SEQUENCE   304 AA;  34146 MW;  20AB91D8BB137A76 CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 304;
  Best Local Similarity   87.5%;  Pred. No. 37;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 AEPEAEPI 10
           |||||||:
Db      80 AEPEAEPL 87


RESULT   9
Q9NAA2
ID   Q9NAA2      PRELIMINARY;      PRT;   340 AA.
AC   Q9NAA2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Y53F4B.17 PROTEIN.
GN   Y53F4B.17.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smye R.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL132949; CAB70107.1; -.
SQ   SEQUENCE   340 AA;  36879 MW;  2ECD1E30582C9E3A CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 340;
  Best Local Similarity   87.5%;  Pred. No. 42;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 AEPEAEPI 10
           |||||||:
Db     247 AEPEAEPL 254


RESULT  10
O96979
ID   O96979      PRELIMINARY;      PRT;   408 AA.
AC   O96979;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TRANSLATION ELONGATION FACTOR 1-ALPHA (FRAGMENT).
GN   TEF1.
```

```
OS   Paranophrys carnivora.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Scuticociliatida;
OC   Paranophryidae; Paranophrys.
OX   NCBI_TaxID=85900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99152616; PubMed=10028290;
RA   Moreira D., Le Guyader H., Philippe H.;
RT   "Unusually high evolutionary rate of the elongation factor 1 alpha
RT   genes from the Ciliophora and its impact on the phylogeny of
RT   eukaryotes.";
RL   Mol. Biol. Evol. 16:234-245(1999).
CC   -!- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
DR   EMBL; AF056103; AAD03258.1; -.
DR   HSSP; Q01698; 1TTT.
DR   InterPro; IPR000795; GTP_EFTU.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; GTP-binding; Protein biosynthesis.
FT   NON_TER       1      1
FT   NON_TER     408    408
SQ   SEQUENCE   408 AA;  45066 MW;  384973BB7F3F5FA1 CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 408;
  Best Local Similarity   77.8%;  Pred. No. 50;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEP 9
           ||: |||||
Db     269 HESLPEAEP 277


RESULT  11
Q9W4Y3
ID   Q9W4Y3      PRELIMINARY;      PRT;   417 AA.
AC   Q9W4Y3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG12497 PROTEIN.
GN   EG:BACR25B3.2 OR CG12497.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC       (LDLRA) DOMAIN.
DR   EMBL; AE003424; AAF45787.1; -.
DR   HSSP; Q07954; 1CR8.
DR   FlyBase; FBgn0040379; EG:BACR25B3.2.
DR   InterPro; IPR002172; LDL_recept_A.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00192; LDLa; 2.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 2.
KW   Glycoprotein.
SQ   SEQUENCE   417 AA;  48624 MW;  2F1CA7F440D1DD01 CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 417;
  Best Local Similarity   87.5%;  Pred. No. 51;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
           |:||||||
Db     382 ESEPEAEP 389


RESULT  12
Q9NEF8
ID   Q9NEF8      PRELIMINARY;      PRT;   435 AA.
AC   Q9NEF8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:BACR25B3.2 PROTEIN.
GN   EG:BACR25B3.2 OR CG12497.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murphy L., Harris D., Barrell B.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC       (LDLRA) DOMAIN.
DR   EMBL; AL138972; CAB72287.1; -.
DR   FlyBase; FBgn0040379; EG:BACR25B3.2.
DR   InterPro; IPR002172; LDL_recept_A.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00192; LDLa; 2.
DR   PROSITE; PS01209; LDLRA_1; 1.
```

```
DR   PROSITE; PS50068; LDLRA_2; 2.
KW   Glycoprotein.
SQ   SEQUENCE   435 AA;  50714 MW;  19D931A6521C3EA1 CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 435;
  Best Local Similarity   87.5%;  Pred. No. 54;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
           |:||||||
Db     400 ESEPEAEP 407


RESULT  13
Q21489
ID   Q21489      PRELIMINARY;   PRT;   848 AA.
AC   Q21489;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   M03C11.2 PROTEIN.
GN   M03C11.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mcmurray A.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z49128; CAA88959.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   SMART; SM00488; DEXDc2; 1.
DR   SMART; SM00491; HELICc2; 1.
SQ   SEQUENCE   848 AA;  96811 MW;  4D8574A99F63734B CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 848;
  Best Local Similarity   77.8%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPI 10
           | ||||||:
Db     506 EPEPEAEPL 514


RESULT  14
Q9W475
ID   Q9W475      PRELIMINARY;      PRT;  1192 AA.
AC   Q9W475;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG3108 PROTEIN.
GN   CG3108.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003435; AAF46083.1; -.
DR   HSSP; P00730; 2CTC.
DR   FlyBase; FBgn0029807; CG3108.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
SQ   SEQUENCE   1192 AA;  132088 MW;  72408D5C5D3D718E CRC64;


  Query Match             66.1%;  Score 39;  DB 5;  Length 1192;
  Best Local Similarity   87.5%;  Pred. No. 1.5e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
           ||:|||||
Db     330 EAQPEAEP 337


RESULT  15
Q08547
ID   Q08547      PRELIMINARY;      PRT;  1277 AA.
AC   Q08547;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
```

DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP4
DE (IMMEDIATE-EARLY PROTEIN IE62).
GN 62.
OS Cercopithecine herpesvirus 9 (simian varicella virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=35246;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94025599; PubMed=8212583;
RA Clarke P., Brunschwig A., Gilden D.H.;
RT "DNA sequence of a simian varicella virus gene that encodes a
RT homologue of varicella zoster virus IE62 and herpes simplex virus
RT ICP4.";
RL Virology 197:45-52(1993).
CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE OF
CC MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING OTHER
CC VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY OF PROTEINS.
DR EMBL; L20759; AAA03549.1; -.
DR InterPro; IPR000923; Copper_blue1.
DR PROSITE; PS00196; COPPER_BLUE; UNKNOWN_1.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 1277 AA; 136978 MW; BB92AAA8C4DCBD9D CRC64;

Query Match 66.1%; Score 39; DB 12; Length 1277;
Best Local Similarity 70.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EAEPEAEPIM 11
| :|||||:
Db 1106 EPDPEAEPII 1115

Search completed: January 3, 2002, 08:47:23
Job time: 411 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:41:02 ; Search time 18.1 Seconds
(without alignments)
22.282 Million cell updates/sec

Title: US-09-444-281-27
Perfect score: 59
Sequence: 1 HEAEPEAEPIM 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	71.2	144	1	AP22_APIME	P35581 apis mellif
2	42	71.2	283	1	AP73_APIME	Q06602 apis mellif
3	40	67.8	129	1	CGB1_CRILO	Q08301 cricetulus
4	39	66.1	486	1	HS1_HUMAN	P14317 homo sapien
5	38	64.4	25	1	NP4_HUMAN	P18078 homo sapien
6	38	64.4	168	1	AP14_APIME	Q06601 apis mellif
7	38	64.4	248	1	GRL1_RAT	Q06605 rattus norv
8	38	64.4	256	1	PRN3_HUMAN	P24158 homo sapien
9	38	64.4	332	1	HS30_YEAST	P25619 saccharomyc
10	37	62.7	226	1	DDN1_BOVIN	P80219 bos taurus
11	37	62.7	245	1	MCT1_SHEEP	P80931 ovis aries
12	37	62.7	246	1	GRAH_HUMAN	P20718 homo sapien
13	37	62.7	247	1	GRAB_HUMAN	P10144 h granzyme
14	37	62.7	247	1	TONB_SERMA	P26185 serratia ma
15	37	62.7	248	1	NKP1_RAT	P18291 rattus norv
16	37	62.7	251	1	MCT3_SHEEP	O46683 ovis aries
17	37	62.7	269	1	YA53_ARCFU	O29209 archaeoglob
18	37	62.7	376	1	XYNA_BACOV	P49942 bacteroides
19	37	62.7	483	1	PHR_SYNLE	P05327 synechococc
20	37	62.7	601	1	PAGT_CAEEL	P34678 caenorhabdi
21	36	61.0	182	1	YSM2_CAEEL	Q10122 caenorhabdi
22	36	61.0	239	1	TONB_ECOLI	P02929 escherichia
23	36	61.0	242	1	TONB_SALTY	P25945 salmonella
24	36	61.0	255	1	TONB_YEREN	Q05740 yersinia en
25	36	61.0	277	1	PS12_ARATH	O23712 arabidopsis
26	36	61.0	372	1	MK11_HUMAN	Q15759 homo sapien
27	36	61.0	423	1	CGB1_RAT	P30277 rattus norv
28	36	61.0	565	1	MOT8_MOUSE	O70324 mus musculu
29	36	61.0	1195	1	KDGD_HUMAN	Q16760 homo sapien
30	35	59.3	171	1	ZUR_ECOLI	P32692 escherichia
31	35	59.3	292	1	SUB1_YEAST	P54000 saccharomyc
32	35	59.3	330	1	XYN2_BACST	P45703 bacillus st
33	35	59.3	433	1	CGB1_HUMAN	P14635 homo sapien
34	35	59.3	465	1	G3BP_MOUSE	P97855 mus musculu
35	35	59.3	466	1	G3BP_HUMAN	Q13283 homo sapien
36	35	59.3	474	1	SY65_DROME	P21521 drosophila
37	35	59.3	486	1	HS1_MOUSE	P49710 mus musculu
38	35	59.3	613	1	MOT8_HUMAN	P36021 homo sapien
39	35	59.3	754	1	RAD4_YEAST	P14736 saccharomyc
40	35	59.3	788	1	TRS1_HCMVA	P09695 human cytom
41	35	59.3	2749	1	IP3R_MOUSE	P11881 mus musculu
42	35	59.3	2749	1	IP3R_RAT	P29994 rattus norv
43	34	57.6	202	1	B3G1_MOUSE	Q9cw73 m galactosy
44	34	57.6	302	1	RS3_HALHA	P15009 halobacteri
45	34	57.6	334	1	B3G1_RAT	O35789 r galactosy

ALIGNMENTS

```
RESULT   1
AP22_APIME
ID   AP22_APIME     STANDARD;      PRT;   144 AA.
AC   P35581; P11525; P11526;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   APIDAECIN PRECURSOR, TYPE 22.
OS   Apis mellifera (Honeybee).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC   Apoidea; Apidae; Apis.
OX   NCBI_TaxID=7460;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93223697; PubMed=8467807;
RA   Casteels-Josson K., Capaci T., Casteels P., Tempst P.;
RT   "Apidaecin multipeptide precursor structure: a putative mechanism for
RT   amplification of the insect antibacterial response.";
RL   EMBO J. 12:1569-1578(1993).
RN   [2]
RP   SEQUENCE (APIDAECINS IA/IB).
RC   TISSUE=Hemolymph;
RX   MEDLINE=90005446; PubMed=2676519;
RA   Casteels P., Ampe C., Jacobs F., Vaeck M., Tempst P.;
RT   "Apidaecins: antibacterial peptides from honeybees.";
RL   EMBO J. 8:2387-2391(1989).
CC   -!- FUNCTION: APIDAECINS HAVE BACTERICIDAL ACTIVITY; PREDOMINANTLY
CC       AGAINST GRAM-NEGATIVE BACTERIA. THEY SEEM TO INTERFERE WITH CELL
CC       PROPAGATION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X72576; CAA51168.1; -.
DR   PIR; S05383; S05383.
DR   PIR; S06675; S06675.
DR   PIR; S35331; S35331.
DR   InterPro; IPR001979; Apidaecin.
DR   Pfam; PF00807; Apidaecin; 4.
KW   Insect immunity; Antibiotic; Hemolymph; Signal; Multigene family;
KW   Cleavage on pair of basic residues; Repeat.
FT   SIGNAL        1     19       POTENTIAL.
FT   PROPEP       35     42
FT   PEPTIDE      43     60       APIDAECIN IB.
FT   PROPEP       63     70
FT   PEPTIDE      71     88       APIDAECIN IB.
FT   PROPEP       91     98
FT   PEPTIDE      99    116       APIDAECIN IB.
FT   PROPEP      119    126
FT   PEPTIDE     127    144       APIDAECIN IA.
```

```
SQ   SEQUENCE   144 AA;  16539 MW;  6FA1AD74CB77108D CRC64;


  Query Match             71.2%;  Score 42;  DB 1;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 EAEPEAEP 9
          ||||||||
Db     35 EAEPEAEP 42


RESULT   2
AP73_APIME
ID   AP73_APIME     STANDARD;      PRT;   283 AA.
AC   Q06602; P11525; P11526;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   APIDAECIN PRECURSOR, TYPE 73 (FRAGMENT).
GN   APID73.
OS   Apis mellifera (Honeybee).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC   Apoidea; Apidae; Apis.
OX   NCBI_TaxID=7460;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93223697; PubMed=8467807;
RA   Casteels-Josson K., Capaci T., Casteels P., Tempst P.;
RT   "Apidaecin multipeptide precursor structure: a putative mechanism for
RT   amplification of the insect antibacterial response.";
RL   EMBO J. 12:1569-1578(1993).
RN   [2]
RP   SEQUENCE OF APIDAECINS IA/IB.
RC   TISSUE=Hemolymph;
RX   MEDLINE=90005446; PubMed=2676519;
RA   Casteels P., Ampe C., Jacobs F., Vaeck M., Tempst P.;
RT   "Apidaecins: antibacterial peptides from honeybees.";
RL   EMBO J. 8:2387-2391(1989).
CC   -!- FUNCTION: APIDAECINS HAVE BACTERICIDAL ACTIVITY; PREDOMINANTLY
CC       AGAINST GRAM-NEGATIVE BACTERIA. THEY SEEM TO INTERFERE WITH CELL
CC       PROPAGATION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72577; CAA51169.1; -.
DR   PIR; S05383; S05383.
DR   PIR; S06675; S06675.
DR   PIR; S35332; S35332.
DR   InterPro; IPR001979; Apidaecin.
DR   Pfam; PF00807; Apidaecin; 9.
DR   ProDom; PD153432; Apidaecin; 2.
KW   Insect immunity; Antibiotic; Hemolymph; Signal; Multigene family;
KW   Cleavage on pair of basic residues; Repeat.
FT   NON_TER       1      1
FT   SIGNAL       <1     18       POTENTIAL.
FT   PROPEP        34     41
FT   PEPTIDE       42     59       APIDAECIN IB.
FT   PROPEP        62     69
FT   PEPTIDE       70     87       APIDAECIN IB.
FT   PROPEP        90     97
FT   PEPTIDE       98    115       APIDAECIN.
FT   PROPEP       118    125
FT   PEPTIDE      126    143       APIDAECIN IB.
FT   PROPEP       146    153
FT   PEPTIDE      154    171       APIDAECIN.
FT   PROPEP       174    182
FT   PEPTIDE      183    199       APIDAECIN IB.
FT   PROPEP       202    209
FT   PEPTIDE      210    227       APIDAECIN IB.
FT   PROPEP       230    237
FT   PEPTIDE      238    255       APIDAECIN IB.
FT   PROPEP       258    265
FT   PEPTIDE      266    283       APIDAECIN IA.
SQ   SEQUENCE   283 AA;  32695 MW;  4EA5FEDECD5E142B CRC64;


  Query Match             71.2%;  Score 42;  DB 1;  Length 283;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 EAEPEAEP 9
          ||||||||
Db     62 EAEPEAEP 69


RESULT   3
CGB1_CRILO
ID   CGB1_CRILO     STANDARD;      PRT;   429 AA.
AC   Q08301;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   G2/MITOTIC-SPECIFIC CYCLIN B1.
GN   CCNB1.
OS   Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10030;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=94095439; PubMed=8270434;
RA   Markiewicz D.A., McKenna W.G., Flick M.B., Maity A., Muschel R.J.;
RT   "The effects of radiation on the expression of a newly cloned and
RT   characterized rat cyclin B mRNA.";
RL   Int. J. Radiat. Oncol. Biol. Phys. 28:135-144(1994).
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC       (MITOSIS) TRANSITION.
CC   -!- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC       SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC       MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC       SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64588; CAA45876.1; -.
DR   PIR; S34224; S34224.
DR   HSSP; P20248; 1JST.
DR   InterPro; IPR000553; Cyclin.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis.
FT   DOMAIN        51     86       LYS-RICH.
SQ   SEQUENCE   429 AA;  48062 MW;  6E0BAE7511A678B7 CRC64;


  Query Match             67.8%;  Score 40;  DB 1;  Length 429;
```

```
  Best Local Similarity   70.0%;  Pred. No. 9.6;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPIM 11
           | ||| ||:|
Db      98 EPEPEPEPVM 107


RESULT   4
HS1_HUMAN
ID   HS1_HUMAN      STANDARD;      PRT;   486 AA.
AC   P14317;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN (HEMATOPOIETIC CELL-
DE   SPECIFIC LYN SUBSTRATE 1) (LCKBP1).
GN   HCLS1 OR HS1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90067934; PubMed=2587259;
RA   Kitamura D., Kaneko H., Miyagoe Y., Ariyasu T., Watanabe T.;
RT   "Isolation and characterization of a novel human gene expressed
RT   specifically in the cells of hematopoietic lineage.";
RL   Nucleic Acids Res. 17:9367-9379(1989).
RN   [2]
RP   SEQUENCE OF 97-108; 193-201 AND 240-248.
RX   MEDLINE=96311348; PubMed=8713105;
RA   Egerton M., Moritz R.L., Druker B., Kelso A., Simpson R.J.;
RT   "Identification of the 70kD heat shock cognate protein (Hsc70) and
RT   alpha-actinin-1 as novel phosphotyrosine-containing proteins in T
RT   lymphocytes.";
RL   Biochem. Biophys. Res. Commun. 224:666-674(1996).
RN   [3]
RP   BINDING TO HAX-1 PROTEIN.
RX   MEDLINE=97211841; PubMed=9058808;
RA   Suzuki Y., Demoliere C., Kitamura D., Takeshita H., Deuschle U.,
RA   Watanabe T.;
RT   "HAX-1, a novel intracellular protein, localized on mitochondria,
RT   directly associates with HS1, a substrate of Src family tyrosine
RT   kinases.";
RL   J. Immunol. 158:2736-2744(1997).
CC   -!- FUNCTION: SUBSTRATE OF THE ANTIGEN RECEPTOR-COUPLED TYROSINE
CC       KINASE. PLAYS A ROLE IN ANTIGEN RECEPTOR SIGNALING FOR BOTH
CC       CLONAL EXPANSION AND DELETION IN LYMPHOID CELLS. DIRECTLY
CC       ASSOCIATES WITH HAX-1, THROUGH BINDING TO ITS C-TERMINAL REGION.
CC       MAY ALSO BE INVOLVED IN THE REGULATION OF GENE EXPRESSION.
CC   -!- SUBUNIT: ASSOCIATES WITH THE SH2 AND SH3 DOMAINS OF LCK.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (PROBABLE).
CC   -!- TISSUE SPECIFICITY: EXPRESSED ONLY IN TISSUES AND CELLS OF
CC       HEMATOPOIETIC ORIGIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN EARLY STAGE OF MYELOID AND
CC       ERYTHROID DIFFERENTIATION.
CC   -!- PTM: PHOSPHORYLATED BY LYN RAPIDLY AFTER CROSSLINKING OF SURFACE
CC       IGM ON B CELLS.
CC   -!- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16663; CAA34651.1; -.
DR   PIR; S07633; S07633.
DR   HSSP; P29355; 1SEM.
DR   MIM; 601306; -.
DR   InterPro; IPR003134; HS1_rep.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF02218; HS1_rep; 4.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Repeat; SH3 domain; Phosphorylation.
FT   DOMAIN       27     66       INVOLVED IN HAX-1 BINDING.
FT   DOMAIN       81    214       3.5 X 37 AA TANDEM REPEATS.
FT   REPEAT       81    116       1.
FT   REPEAT      117    153       2.
FT   REPEAT      154    190       3.
FT   REPEAT      191    214       4 (INCOMPLETE).
FT   DOMAIN      428    486       SH3.
FT   CONFLICT    241    242       KF -> FK (IN REF. 2).
SQ   SEQUENCE   486 AA;  53998 MW;  61AE637157DF5DF2 CRC64;


  Query Match             66.1%;  Score 39;  DB 1;  Length 486;
  Best Local Similarity   77.8%;  Pred. No. 16;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEP 9
           :||||| ||
Db     360 YEAEPEPEP 368


RESULT   5
NP4_HUMAN
ID   NP4_HUMAN      STANDARD;      PRT;    25 AA.
AC   P18078;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUTROPHIL PROTEINASE 4 (EC 3.4.21.-) (NP-4) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91025622; PubMed=2121162;
RA   Ohlsson K., Linder C., Rosengren M.;
RT   "Monoclonal antibodies specific for neutrophil proteinase 4.
RT   Production and use for isolation of the enzyme.";
RL   Biol. Chem. Hoppe-Seyler 371:549-555(1990).
CC   -!- TISSUE SPECIFICITY: NEUTROPHILS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
DR   HSSP; P24158; 1FUJ.
DR   MEROPS; S01.134; -.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; PARTIAL.
DR   PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
DR   PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
KW   Hydrolase; Serine protease; Glycoprotein.
FT   NON_TER      25     25
SQ   SEQUENCE   25 AA;  2606 MW;  B1CB2038274575B0 CRC64;


  Query Match             64.4%;  Score 38;  DB 1;  Length 25;
  Best Local Similarity   54.5%;  Pred. No. 1.2;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db       5 HEAQPHSRPYM 15


RESULT   6
```

```
AP14_APIME
ID   AP14_APIME     STANDARD;      PRT;   168 AA.
AC   Q06601; P11525; P11526; P11527;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   APIDAECIN PRECURSOR, TYPE 14.
GN   APID14.
OS   Apis mellifera (Honeybee).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC   Apoidea; Apidae; Apis.
OX   NCBI_TaxID=7460;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93223697; PubMed=8467807;
RA   Casteels-Josson K., Capaci T., Casteels P., Tempst P.;
RT   "Apidaecin multipeptide precursor structure: a putative mechanism for
RT   amplification of the insect antibacterial response.";
RL   EMBO J. 12:1569-1578(1993).
RN   [2]
RP   SEQUENCE OF APIDAECINS IA/IB/II.
RC   TISSUE=Hemolymph;
RX   MEDLINE=90005446; PubMed=2676519;
RA   Casteels P., Ampe C., Jacobs F., Vaeck M., Tempst P.;
RT   "Apidaecins: antibacterial peptides from honeybees.";
RL   EMBO J. 8:2387-2391(1989).
CC   -!- FUNCTION: APIDAECINS HAVE BACTERICIDAL ACTIVITY; PREDOMINANTLY
CC       AGAINST GRAM-NEGATIVE BACTERIA. THEY SEEM TO INTERFERE WITH CELL
CC       PROPAGATION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72575; CAA51167.1; -.
DR   PIR; S05383; S05383.
DR   PIR; S06675; S06675.
DR   PIR; S06676; S06676.
DR   PIR; S35330; S35330.
DR   InterPro; IPR001979; Apidaecin.
DR   Pfam; PF00807; Apidaecin; 5.
DR   ProDom; PD153432; Apidaecin; 1.
KW   Insect immunity; Antibiotic; Hemolymph; Signal; Multigene family;
KW   Cleavage on pair of basic residues; Repeat.
FT   SIGNAL        1     19       POTENTIAL.
FT   PROPEP       35     42
FT   PEPTIDE      43     60       APIDAECIN II.
FT   PROPEP       63     70
FT   PEPTIDE      71     88       APIDAECIN II.
FT   PROPEP       91     98
FT   PEPTIDE      99    116       APIDAECIN IB.
FT   PROPEP      119    124
FT   PEPTIDE     125    142       APIDAECIN IB.
FT   PROPEP      145    150
FT   PEPTIDE     151    168       APIDAECIN IA.
SQ   SEQUENCE   168 AA;  19380 MW;  594B931254C04A37 CRC64;


  Query Match             64.4%;  Score 38;  DB 1;  Length 168;
  Best Local Similarity   87.5%;  Pred. No. 8.3;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
            ||:|||||
Db      35 EAKPEAEP 42


RESULT   7
```

```
GRL1_RAT
ID   GRL1_RAT       STANDARD;      PRT;   248 AA.
AC   Q06605;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   GRANZYME-LIKE PROTEIN I PRECURSOR (EC 3.4.21.-).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Duodenum;
RX   MEDLINE=93285341; PubMed=8508925;
RA   Amerik A.Y., Yarovoi S.V., Grigorenko V.G., Antonov V.K.;
RT   "Identification, sequence analysis, and characterization of cDNA
RT   clones encoding two granzyme-like serine proteinases from rat
RT   duodenum.";
RL   FEBS Lett. 324:226-230(1993).
CC   -!- FUNCTION: THIS ENZYME IS NECESSARY FOR TARGET CELL LYSIS IN
CC       CELL-MEDIATED IMMUNE RESPONSES.
CC   -!- TISSUE SPECIFICITY: DUODENUM.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. STRONGEST TO GRANZYMES AND TO MAST CELL PROTEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66693; CAA47235.1; -.
DR   HSSP; P04187; 2CP1.
DR   MEROPS; S01.136; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein.
FT   SIGNAL        1     18       BY SIMILARITY.
FT   PROPEP       19     20       ACTIVATION PEPTIDE.
FT   CHAIN        21    248       GRANZYME-LIKE PROTEIN I.
FT   ACT_SITE     65     65       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    109    109       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    204    204       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     50     66       BY SIMILARITY.
FT   DISULFID    143    210       BY SIMILARITY.
FT   DISULFID    174    189       BY SIMILARITY.
FT   CARBOHYD     72     72       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   248 AA;  27235 MW;  036C81B43A88F972 CRC64;


  Query Match             64.4%;  Score 38;  DB 1;  Length 248;
  Best Local Similarity   54.5%;  Pred. No. 12;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
            |||:| : | |
Db      25 HEADPHSRPYM 35


RESULT   8
PRN3_HUMAN
ID   PRN3_HUMAN     STANDARD;      PRT;   256 AA.
AC   P24158; P15637;
DT   01-APR-1990 (Rel. 14, Created)
```

```
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MYELOBLASTIN PRECURSOR (EC 3.4.21.76) (LEUKOCYTE PROTEINASE 3) (PR-3)
DE   (PR3) (AGP7) (WEGENER'S AUTOANTIGEN) (P29) (C-ANCA ANTIGEN).
GN   PRTN3 OR MBN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92021028; PubMed=1681549;
RA   Labbaye C., Musette P., Cayre Y.E.;
RT   "Wegener autoantigen and myeloblastin are encoded by a single mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9253-9256(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K.J., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 2-256 FROM N.A., AND SEQUENCE OF 48-71 AND 156-181.
RX   MEDLINE=91079774; PubMed=2258701;
RA   Campanelli D., Melchior M., Fu Y., Nakata M., Shuman H., Nathan C.,
RA   Gabay J.E.;
RT   "Cloning of cDNA for proteinase 3: a serine protease, antibiotic, and
RT   autoantigen from human neutrophils.";
RL   J. Exp. Med. 172:1709-1715(1990).
RN   [4]
RP   SEQUENCE OF 1-20 AND 22-256 FROM N.A.
RX   MEDLINE=92390417; PubMed=1518849;
RA   Zimmer M., Medcalf R.L., Fink T.M., Mattmann C., Lichter P.,
RA   Jenne D.E.;
RT   "Three human elastase-like genes coordinately expressed in the
RT   myelomonocyte lineage are organized as a single genetic locus on
RT   19pter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:8215-8219(1992).
RN   [5]
RP   SEQUENCE OF 42-256 FROM N.A.
RX   MEDLINE=90090622; PubMed=2598267;
RA   Bories D., Raynal M.-C., Solomon D.H., Darzynkiewicz Z., Cayre Y.E.;
RT   "Down-regulation of a serine protease, myeloblastin, causes growth
RT   arrest and differentiation of promyelocytic leukemia cells.";
RL   Cell 59:959-968(1989).
RN   [6]
RP   SEQUENCE OF 28-67 AND 228-244.
RX   MEDLINE=91236723; PubMed=2033050;
RA   Rao N.V., Wehner N.G., Marshall B.C., Gray W.R., Gray B.H.,
RA   Hoidal J.R.;
RT   "Characterization of proteinase-3 (PR-3), a neutrophil serine
RT   proteinase. Structural and functional properties.";
RL   J. Biol. Chem. 266:9540-9548(1991).
RN   [7]
RP   SEQUENCE OF 28-47.
RX   MEDLINE=89315847; PubMed=2501794;
RA   Gabay J.E., Scott R.W., Campanelli D., Griffith J., Wilde C.,
RA   Marra M.N., Seeger M., Nathan C.F.;
RT   "Antibiotic proteins of human polymorphonuclear leukocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:5610-5614(1989).
RN   [8]
RP   SEQUENCE OF 28-47 AND 196-219.
RX   MEDLINE=90130450; PubMed=2404977;
RA   Wilde C.G., Snable J.L., Griffith J.E., Scott R.W.;
RT   "Characterization of two azurphil granule proteases with active-site
RT   homology to neutrophil elastase.";
RL   J. Biol. Chem. 265:2038-2041(1990).
RN   [9]
RP   SEQUENCE OF 28-48, AND IDENTITY OF WEGENER'S AUTOANTIGEN WITH PR-3.
RX   MEDLINE=90332035; PubMed=2377228;
RA   Jenne D.E., Tschopp J., Luedemann J., Utecht B., Gross W.L.;
RT   "Wegener's autoantigen decoded.";
RL   Nature 346:520-520(1990).
RN   [10]
RP   IDENTITY OF WEGENER'S AUTOANTIGEN WITH PROTEINASE 3.
RX   MEDLINE=91055123; PubMed=2242436;
RA   Gupta S.K., Niles J.L., McCluskey R.T., Arnaout M.A.;
RT   "Identity of Wegener's autoantigen (p29) with proteinase 3 and
RT   myeloblastin.";
RL   Blood 76:2162-2162(1990).
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX   MEDLINE=96346178; PubMed=8757293;
RA   Fujinaga M., Charnaia M.M., Halenbeck R., Koths K., James M.N.G.;
RT   "The crystal structure of PR3, a neutrophil serine proteinase antigen
RT   of Wegener's granulomatosis antibodies.";
RL   J. Mol. Biol. 261:267-278(1996).
CC   -!- FUNCTION: POLYMORPHONUCLEAR LEUKOCYTE SERINE PROTEASE THAT
CC       DEGRADES ELASTIN, FIBRONECTIN, LAMININ, VITRONECTIN, AND COLLAGEN
CC       TYPES I, III, AND IV (IN VITRO) AND CAUSES EMPHYSEMA WHEN
CC       ADMINISTERED BY TRACHEAL INSUFFLATION TO HAMSTERS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN, BY
CC       PREFERENTIAL CLEAVAGE: ALA-|-XAA > VAL-|-XAA.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. ELASTASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56132; CAA39598.1; -.
DR   EMBL; AC004799; AAC18958.1; -.
DR   EMBL; M75154; AAA59558.1; -.
DR   EMBL; M96839; AAB59493.1; -.
DR   EMBL; M96838; AAB59493.1; JOINED.
DR   EMBL; M96837; AAB59493.1; JOINED.
DR   EMBL; X55668; CAA39203.1; -.
DR   EMBL; M29142; AAA36342.1; -.
DR   EMBL; M96628; AAB59364.1; -.
DR   PIR; A43983; PRHU3.
DR   PDB; 1FUJ; 11-JUL-96.
DR   MEROPS; S01.134; -.
DR   MIM; 177020; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1     25
FT   PROPEP      26     27
FT   CHAIN       28    248       MYELOBLASTIN.
FT   PROPEP     249    256
FT   ACT_SITE    71     71       CHARGE RELAY SYSTEM.
FT   ACT_SITE   118    118       CHARGE RELAY SYSTEM.
FT   ACT_SITE   203    203       CHARGE RELAY SYSTEM.
FT   CARBOHYD   129    129       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   174    174       N-LINKED (GLCNAC. . .).
FT   DISULFID    56     72
FT   DISULFID   152    209
FT   DISULFID   182    188
FT   DISULFID   199    224
FT   CONFLICT     2      2       A -> R (IN REF. 3).
```

```
FT   CONFLICT     46     46       Q -> E (IN REF. 7 AND 8).
FT   CONFLICT     64     64       S -> D (IN REF. 6).
FT   CONFLICT     70     70       A -> P (IN REF. 1).
FT   CONFLICT    119    119       V -> I (IN REF. 1 AND 5).
FT   CONFLICT    135    136       AT -> TS (IN REF. 1 AND 5).
FT   CONFLICT    255    255       MISSING (IN REF. 3).
SQ   SEQUENCE   256 AA;  27807 MW;  CBECA36D8C4B2A40 CRC64;


  Query Match             64.4%;  Score 38;  DB 1;  Length 256;
  Best Local Similarity   54.5%;  Pred. No. 13;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db     32 HEAQPHSRPYM 42


RESULT   9
HS30_YEAST
ID   HS30_YEAST     STANDARD;      PRT;   332 AA.
AC   P25619; Q04556;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   30 KDA HEAT SHOCK PROTEIN.
GN   HSP30 OR YCR021C OR YCR21C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SKQ2N;
RX   MEDLINE=93306747; PubMed=8319300;
RA   Regnacq M., Boucherie H.;
RT   "Isolation and sequence of HSP30, a yeast heat-shock gene coding for
RT   a hydrophobic membrane protein.";
RL   Curr. Genet. 23:435-442(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Feldmann H., Mannhaupt G., Vetter I.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLY COOPERATES WITH OTHER HEAT SHOCK PROTEINS IN
CC       THE TRANSLOCATION OF POLYPEPTIDES THROUGH MEMBRANES. IT MAY
CC       COUNTERACT THE ALTERING EFFECT OF HEAT SHOCK ON THE PLASMA
CC       MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING THE ENTRY INTO STATIONARY
CC       PHASE RESULTING FROM GLUCOSE LIMITATION.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL OPSIN FAMILY. HSP30
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M93123; AAA02903.1; -.
DR   EMBL; X59720; CAA42313.1; -.
DR   PIR; S30781; S30781.
DR   PIR; S19432; S19432.
DR   SGD; S0000615; HSP30.
KW   Heat shock; Transmembrane.
FT   TRANSMEM     35     55       POTENTIAL.
FT   TRANSMEM     66     86       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    182    202       POTENTIAL.
FT   TRANSMEM    216    236       POTENTIAL.
FT   TRANSMEM    249    269       POTENTIAL.
FT   DOMAIN      290    332       GLU-RICH (ACIDIC).
SQ   SEQUENCE   332 AA;  37044 MW;  260474A481D29AC5 CRC64;


  Query Match             64.4%;  Score 38;  DB 1;  Length 332;
  Best Local Similarity   87.5%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAE 8
          || |||||
Db    318 HEPEPEAE 325


RESULT  10
DDN1_BOVIN
ID   DDN1_BOVIN     STANDARD;      PRT;   226 AA.
AC   P80219;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DUODENASE I (EC 3.4.21.-) (DUODENUM SERINE PROTEASE).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Duodenum;
RX   MEDLINE=95172076; PubMed=7867649;
RA   Zamolodchikova T.S., Vorotyntseva T.I., Nazimov I.V., Grishina G.A.;
RT   "Duodenase, a new serine protease of unusual specificity from bovine
RT   duodenal mucosa. Primary structure of the enzyme.";
RL   Eur. J. Biochem. 227:873-879(1995).
RN   [2]
RP   SEQUENCE OF 1-20 AND 172-183.
RX   MEDLINE=93048618; PubMed=1425193;
RA   Antonov V.K., Vorotyntseva T.I., Zamolodchikova T.S.;
RT   "Duodenase -- a new serine proteinase with unusual specificity.";
RL   Dokl. Akad. Nauk SSSR 324:1318-1322(1992).
RN   [3]
RP   SEQUENCE OF 1-24.
RX   MEDLINE=95172075; PubMed=7867648;
RA   Zamolodchikova T.S., Vorotyntseva T.I., Antonov V.K.;
RT   "Duodenase, a new serine protease of unusual specificity from bovine
RT   duodenal mucosa. Purification and properties.";
RL   Eur. J. Biochem. 227:866-872(1995).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: LYS-, ARG-, TYR-, PHE-,
CC       LEU-.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: THE OPTIMUM PH AND TEMPERATURE OF DUODENASE I ARE
CC       8.0 AND 50 DEGREES CELSIUS, RESPECTIVELY.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.142; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Glycoprotein.
FT   ACT_SITE     44     44       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE     87     87       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    181    181       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     29     45       BY SIMILARITY.
FT   DISULFID    121    187       BY SIMILARITY.
FT   DISULFID    152    166       BY SIMILARITY.
FT   CARBOHYD     50     50       N-LINKED (GLCNAC. . .).
```

```
SQ   SEQUENCE   226 AA;  25051 MW;  A4CA2CE66736CD12 CRC64;


  Query Match           62.7%;  Score 37;  DB 1;  Length 226;
  Best Local Similarity   54.5%;  Pred. No. 17;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db      5 HEAKPHSRPYM 15


RESULT  11
MCT1_SHEEP
ID   MCT1_SHEEP     STANDARD;      PRT;   245 AA.
AC   P80931;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MAST CELL PROTEASE 1A PRECURSOR (EC 3.4.21.-) (SMCP-1A).
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 176-185.
RC   TISSUE=Mast cells;
RX   MEDLINE=98343972; PubMed=9677343;
RA   McAleese S.M., Pemberton A.D., McGrath M.E., Huntley J.F.,
RA   Miller H.R.P.;
RT   "Sheep mast-cell proteinases-1 and -3: cDNA cloning, primary
RT   structure and molecular modelling of the enzymes and further studies
RT   on substrate specificity.";
RL   Biochem. J. 333:801-809(1998).
RN   [2]
RP   SEQUENCE OF 20-44.
RA   Miller H.R.P., Huntley J.F., Newlands G.F.J.;
RL   (In) Caughey G.H. (eds.);
RL   Mast cell proteases in immunology and biology, pp.203-235, Marcel
RL   Dekker, New York (1995).
RN   [3]
RP   SEQUENCE OF 20-29.
RC   TISSUE=Gastric mucosa;
RX   MEDLINE=97184650; PubMed=9032451;
RA   Pemberton A.D., Huntley J.F., Miller H.R.P.;
RT   "Sheep mast cell proteinase-1: characterization as a member of a new
RT   class of dual-specific ruminant chymases.";
RL   Biochem. J. 321:665-670(1997).
CC   -!- FUNCTION: HAS A CHYMOTRYPSIN-LIKE AND TRYPSIN-LIKE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: SECRETORY GRANULES.
CC   -!- TISSUE SPECIFICITY: MUCOSAL MAST CELLS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMES AND TO MAST CELL
CC       PROTEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14654; CAA74984.1; -.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.142; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   PROPEP      18     19       ACTIVATION PEPTIDE.
FT   CHAIN       20    245       MAST CELL PROTEASE 1A.
FT   ACT_SITE    63     63       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   107    107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   201    201       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    48     64       BY SIMILARITY.
FT   DISULFID   141    207       BY SIMILARITY.
FT   DISULFID   172    186       BY SIMILARITY.
FT   CONFLICT    20     20       I -> F (IN REF. 2; AA SEQUENCE).
SQ   SEQUENCE   245 AA;  26877 MW;  C362CF3367FFF180 CRC64;


  Query Match           62.7%;  Score 37;  DB 1;  Length 245;
  Best Local Similarity   54.5%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db     24 HEAKPHSRPYM 34


RESULT  12
GRAH_HUMAN
ID   GRAH_HUMAN     STANDARD;      PRT;   246 AA.
AC   P20718;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GRANZYME H PRECURSOR (EC 3.4.21.-) (CYTOTOXIC T-LYMPHOCYTE PROTEINASE)
DE   (CATHEPSIN G-LIKE 2) (CTSGL2) (CCP-X) (CYTOTOXIC SERINE PROTEASE-C)
DE   (CSP-C).
GN   GZMH OR CTSGL2 OR CGL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90298167; PubMed=2193684;
RA   Meier M., Kwong P.C., Fregeau C.J., Atkinson E.A., Burrington M.,
RA   Ehrman N., Sorensen O., Lin C.C., Wilkins J., Bleackley R.C.;
RT   "Cloning of a gene that encodes a new member of the human cytotoxic
RT   cell protease family.";
RL   Biochemistry 29:4042-4049(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91265452; PubMed=2049336;
RA   Haddad P., Jenne D.E., Tschopp J., Clement M.V., Mathieu-Mahul D.,
RA   Sasportes M.;
RT   "Structure and evolutionary origin of the human granzyme H gene.";
RL   Int. Immunol. 3:57-66(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90385483; PubMed=2402757;
RA   Klein J.L., Selvakumar A., Trapani J.A., Dupont B.;
RT   "Characterization of a novel, human cytotoxic lymphocyte-specific
RT   serine protease cDNA clone (CSP-C).";
RL   Tissue Antigens 35:220-228(1990).
CC   -!- FUNCTION: THIS ENZYME IS PROBABLY NECESSARY FOR TARGET CELL
CC       LYSIS IN CELL-MEDIATED IMMUNE RESPONSES.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC
CC       T-LYMPHOCYTES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMES AND TO MAST CELL
CC       PROTEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; J02907; AAA76859.1; -.
DR   EMBL; M57888; AAA03514.1; -.
DR   EMBL; M36118; AAA03248.1; -.
DR   EMBL; M72150; AAA74885.1; -.
DR   PIR; A32692; A32692.
DR   HSSP; P04187; 2CP1.
DR   MEROPS; S01.147; -.
DR   MIM; 116831; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; T-cell; Cytolysis.
FT   SIGNAL       1     18
FT   PROPEP      19     20       ACTIVATION PEPTIDE.
FT   CHAIN       21    246       GRANZYME H.
FT   ACT_SITE    64     64       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   108    108       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   202    202       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    49     65       BY SIMILARITY.
FT   DISULFID   142    208       BY SIMILARITY.
FT   DISULFID   172    187       BY SIMILARITY.
FT   CARBOHYD    71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   179    179       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   246 AA;  27315 MW;  CA6A87F3DA5F1E71 CRC64;


  Query Match             62.7%;  Score 37;  DB 1;  Length 246;
  Best Local Similarity   54.5%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      25 HEAKPHSRPYM 35


RESULT  13
GRAB_HUMAN
ID   GRAB_HUMAN      STANDARD;      PRT;   247 AA.
AC   P10144;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GRANZYME B PRECURSOR (EC 3.4.21.79) (T-CELL SERINE PROTEASE 1-3E)
DE   (CYTOTOXIC T-LYMPHOCYTE PROTEINASE 2) (LYMPHOCYTE PROTEASE) (SECT)
DE   (GRANZYME 2) (CATHEPSIN G-LIKE 1) (CTSGL1) (CTLA-1) (FRAGMENTIN 2)
DE   (HUMAN LYMPHOCYTE PROTEIN) (HLP) (C11).
GN   GZMB OR CTLA1 OR GRB OR CSPB OR CGL1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87224164; PubMed=2953813;
RA   Schmid J., Weissmann C.;
RT   "Induction of mRNA for a serine protease and a
RT   beta-thromboglobulin-like protein in mitogen-stimulated human
RT   leukocytes.";
RL   J. Immunol. 139:250-256(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88198184; PubMed=3258865;
RA   Caputo A., Fahey D., Lloyd C., Vozab R., McCairns E., Rowe P.B.;
RT   "Structure and differential mechanisms of regulation of expression of
RT   a serine esterase gene in activated human T lymphocytes.";
RL   J. Biol. Chem. 263:6363-6369(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88320548; PubMed=3261871;
RA   Trapani J.A., Klein J.L., White P.C., Dupont B.;
RT   "Molecular cloning of an inducible serine esterase gene from human
RT   cytotoxic lymphocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6924-6928(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89357968; PubMed=2788607;
RA   Klein J.L., Shows T.B., Dupont B., Trapani J.A.;
RT   "Genomic organization and chromosomal assignment for a serine
RT   protease gene (CSPB) expressed by human cytotoxic lymphocytes.";
RL   Genomics 5:110-117(1989).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90308320; PubMed=2365998;
RA   Caputo A., Sauer D.E., Rowe P.B.;
RT   "Nucleotide sequence and genomic organization of a human T lymphocyte
RT   serine protease gene.";
RL   J. Immunol. 145:737-744(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90236319; PubMed=2332171;
RA   Haddad P., Clement M.V., Bernard O., Larsen C.J., Degos L.,
RA   Sasportes M., Mathieu-Mahul D.;
RT   "Structural organization of the hCTLA-1 gene encoding human granzyme
RT   B.";
RL   Gene 87:265-271(1990).
RN   [7]
RP   SEQUENCE OF 21-40, AND CHARACTERIZATION.
RX   MEDLINE=89009866; PubMed=3262682;
RA   Hameed A., Lowrey D.M., Lichtenheld M., Podack E.R.;
RT   "Characterization of three serine esterases isolated from human IL-2
RT   activated killer cells.";
RL   J. Immunol. 141:3142-3147(1988).
RN   [8]
RP   SEQUENCE OF 21-40, AND CHARACTERIZATION.
RX   MEDLINE=89035468; PubMed=3263427;
RA   Kraehenbuhl O., Rey C., Jenne D.E., Lanzavecchia A., Groscurth P.,
RA   Carrel S., Tschopp J.;
RT   "Characterization of granzymes A and B isolated from granules of
RT   cloned human cytotoxic T lymphocytes.";
RL   J. Immunol. 141:3471-3477(1988).
RN   [9]
RP   SEQUENCE OF 21-38.
RX   MEDLINE=91093203; PubMed=1985927;
RA   Poe M., Blake J.T., Boulton D.A., Gammon M., Sigal N.H., Wu J.K.,
RA   Zweerink H.J.;
RT   "Human cytotoxic lymphocyte granzyme B. Its purification from
RT   granules and the characterization of substrate and inhibitor
RT   specificity.";
RL   J. Biol. Chem. 266:98-103(1991).
CC   -!- FUNCTION: THIS ENZYME IS NECESSARY FOR TARGET CELL LYSIS IN CELL-
CC       MEDIATED IMMUNE RESPONSES. IT CLEAVES AFTER ASP. SEEMS TO BE
CC       LINKED TO AN ACTIVATION CASCADE OF CASPASES (ASPARTATE-SPECIFIC
CC       CYSTEINE PROTEASES) RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES
CC       CASPASE-3, -7, -9 AND 10 TO GIVE RISE TO ACTIVE ENZYMES MEDIATING
CC       APOPTOSIS.
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-|-XAA >> ASN-|-XAA
CC       > MET-|-XAA, SER-|-XAA.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC T-
CC       LYMPHOCYTES AND NATURAL KILLER CELLS.
CC   -!- INDUCTION: BY STAPHYLOCOCCAL ENTEROTOXIN A (SEA) IN PERIPHERAL
CC       BLOOD LEUKOCYTES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMES AND TO MAST CELL
CC       PROTEASES.
CC   ---------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-slb.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M17016; AAA36627.1; -.
DR   EMBL; J03189; AAA36603.1; -.
DR   EMBL; J04071; AAA52118.1; -.
DR   EMBL; J03072; AAB59528.1; -.
DR   EMBL; M38193; AAA67124.1; -.
DR   EMBL; M28879; AAA75490.1; -.
DR   PIR; A28659; A28659.
DR   PIR; A31405; A31405.
DR   PIR; A32168; A32168.
DR   PIR; B30525; B30525.
DR   PIR; B30526; B30526.
DR   HSSP; P04187; 2CP1.
DR   MEROPS; S01.010; -.
DR   MIM; 123910; -.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
KW   Apoptosis.
FT   SIGNAL       1     18
FT   PROPEP      19     20       ACTIVATION PEPTIDE.
FT   CHAIN       21    247       GRANZYME B.
FT   ACT_SITE    64     64       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   108    108       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   203    203       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    49     65       BY SIMILARITY.
FT   DISULFID   142    209       BY SIMILARITY.
FT   DISULFID   173    188       BY SIMILARITY.
FT   CARBOHYD    71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    55     55       Q -> R (IN REF. 2).
FT   CONFLICT    72     72       V -> G (IN REF. 3).
FT   CONFLICT    94     94       P -> A (IN REF. 2 AND 6).
FT   CONFLICT   212    212       V -> C (IN REF. 4).
SQ   SEQUENCE   247 AA;  27688 MW;  684FF605D6F2F4FB CRC64;


  Query Match             62.7%;  Score 37;  DB 1;  Length 247;
  Best Local Similarity   54.5%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      25 HEAKPHSRPYM 35


RESULT  14
TONB_SERMA
ID   TONB_SERMA     STANDARD;      PRT;   247 AA.
AC   P26185;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   TONB PROTEIN.
GN   TONB.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
OX   NCBI_TaxID=615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W225;
RX   MEDLINE=92140042; PubMed=1838128;
RA   Gaisser S., Braun V.;
RT   "The tonB gene of Serratia marcescens: sequence, activity and partial
RT   complementation of Escherichia coli tonB mutants.";
RL   Mol. Microbiol. 5:2777-2787(1991).
CC   -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC       CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC       THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES SUCH AS COBALAMIN,
CC       AND VARIOUS IRON COMPOUNDS (SUCH AS IRON DICITRATE, ENTEROCHELIN,
CC       AEROBACTIN, ETC.). IN THE ABSENCE OF TONB THESE RECEPTORS BIND
CC       THEIR SUBSTRATES BUT DO NOT CARRY OUT ACTIVE TRANSPORT. TONB ALSO
CC       INTERACTS WITH SOME COLICINS AND IS INVOLVED IN THE ENERGY-
CC       DEPENDENT, IRREVERSIBLE STEPS OF BACTERIOPHAGES PHI-80 AND T1
CC       INFECTION. IT COULD ACT TO TRANSDUCE ENERGY FROM THE CYTOPLASMIC
CC       MEMBRANE TO SPECIFIC ENERGY-REQUIRING PROCESSES IN THE OUTER
CC       MEMBRANE, RESULTING IN THE RELEASE INTO THE PERIPLASM OF LIGANDS
CC       BOUND BY THESE OUTER MEMBRANE PROTEINS (BY SIMILARITY).
CC   -!- SUBUNIT: THE ACCESSORY PROTEINS EXBB AND EXBD SEEM TO FORM A
CC       COMPLEX WITH TONB.
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
CC       MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC       PERIPLASM.
CC   -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X60996; CAA43308.1; -.
DR   InterPro; IPR003538; TonB.
KW   Transport; Protein transport; Bacteriocin transport; Inner membrane;
KW   Periplasmic; Transmembrane; Signal-anchor; Repeat; Phage recognition.
FT   DOMAIN       1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    12     35       SIGNAL-ANCHOR (POTENTIAL).
FT   DOMAIN      36    247       PERIPLASMIC (POTENTIAL).
FT   DOMAIN      76     85       5 X 2 AA TANDEM REPEATS OF E-P.
FT   DOMAIN     101    110       5 X 2 AA TANDEM REPEATS OF K-P.
SQ   SEQUENCE   247 AA;  27389 MW;  46EE6E869ED8B64B CRC64;


  Query Match             62.7%;  Score 37;  DB 1;  Length 247;
  Best Local Similarity   70.0%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPIM 11
           | ||| |||:
Db      78 EPEPEPEPIV 87


RESULT  15
NKP1_RAT
ID   NKP1_RAT       STANDARD;      PRT;   248 AA.
AC   P18291;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NATURAL KILLER CELL PROTEASE 1 PRECURSOR (EC 3.4.21.-) (RNKP-1)
DE   (FRAGMENTIN).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=90171602; PubMed=2307850;
RA   Zunino S.J., Bleackley R.C., Martinez J., Hudig D.;
RT   "RNKP-1, a novel natural killer-associated serine protease gene
RT   cloned from RNK-16 cytotoxic lymphocytes.";
```

```
RL   J. Immunol. 144:2001-2009(1990).
RN   [2]
RP   SEQUENCE OF 21-53.
RX   MEDLINE=92091788; PubMed=1727874;
RA   Sayers T.J., Wiltrout T.A., Sowder R., Munger W.L., Smyth M.J.,
RA   Henderson L.E.;
RT   "Purification of a factor from the granules of a rat natural killer
RT   cell line (RNK) that reduces tumor cell growth and changes tumor
RT   morphology. Molecular identity with a granule serine protease
RT   (RNKP-1).";
RL   J. Immunol. 148:292-300(1992).
RN   [3]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=92121838; PubMed=1732416;
RA   Shi L., Kraut R.P., Aebersold R., Greenberg A.H.;
RT   "A natural killer cell granule protein that induces DNA fragmentation
RT   and apoptosis.";
RL   J. Exp. Med. 175:553-566(1992).
CC   -!- FUNCTION: THIS ENZYME IS PROBABLY NECESSARY FOR TARGET CELL
CC       LYSIS IN CELL-MEDIATED IMMUNE RESPONSES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMES AND TO MAST CELL
CC       PROTEASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M34097; AAA42055.1; -.
DR   PIR; A43520; A43520.
DR   HSSP; P04187; 2CP1.
DR   MEROPS; S01.136; -.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; T-cell; Cytolysis.
FT   SIGNAL        1     18
FT   PROPEP       19     20       ACTIVATION PEPTIDE.
FT   CHAIN        21    248       NATURAL KILLER CELL PROTEASE 1.
FT   ACT_SITE     65     65       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    109    109       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    204    204       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     50     66       BY SIMILARITY.
FT   DISULFID    143    210       BY SIMILARITY.
FT   DISULFID    174    189       BY SIMILARITY.
FT   CONFLICT     98     98       H -> F (IN REF. 3).
FT   CONFLICT    138    138       K -> D (IN REF. 3).
SQ   SEQUENCE   248 AA;  27326 MW;  6F52089DDACCC88C CRC64;


  Query Match             62.7%;  Score 37;  DB 1;  Length 248;
  Best Local Similarity   54.5%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      25 HEAKPHSRPYM 35


Search completed: January  4, 2002, 08:47:47
Job time: 405 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:38:42 ; Search time 27.18 Seconds
(without alignments)
30.829 Million cell updates/sec

Title: US-09-444-281-27
Perfect score: 59
Sequence: 1 HEAEPEAEPIM 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	71.2	144	2	S35331	apidaecin 22 precu
2	42	71.2	283	2	S35332	apidaecin 73 precu
3	40	67.8	429	2	S34224	cyclin B - long-ta
4	39	66.1	251	2	T39332	hypothetical prote
5	39	66.1	486	2	S07633	hematopoietic line
6	39	66.1	848	2	T23694	hypothetical prote
7	39	66.1	1593	2	T22028	hypothetical prote
8	38	64.4	150	2	C72724	hypothetical prote
9	38	64.4	168	2	S35330	apidaecin 14 precu
10	38	64.4	248	2	S33755	granzyme-like prot
11	38	64.4	256	1	PRHU3	proteinase 3 (EC 3
12	38	64.4	288	2	T51059	hypothetical prote
13	38	64.4	332	2	S31848	heat shock protein
14	37	62.7	21	2	S69371	duodenase - bovine
15	37	62.7	125	2	T36257	hypothetical prote
16	37	62.7	175	2	T05669	hypothetical prote
17	37	62.7	226	2	S69370	duodenase - bovine
18	37	62.7	246	2	A32692	cytotoxic T-lympho
19	37	62.7	247	2	S18592	tonB protein - Ser
20	37	62.7	248	2	S43259	granzyme-like prot
21	37	62.7	248	2	A43520	natural killer cel
22	37	62.7	251	2	T10262	mast cell serine p
23	37	62.7	265	2	B81138	phosphatidylserine
24	37	62.7	265	2	C81883	probable membrane
25	37	62.7	269	2	E69381	hypothetical prote
26	37	62.7	281	1	A61021	granzyme B (EC 3.4
27	37	62.7	356	2	T00881	probable PCF2-like
28	37	62.7	376	2	S55892	endo-1,4-beta-xyla
29	37	62.7	484	2	S00757	deoxyribodipyrimid
30	37	62.7	612	2	T42243	probable polypepti
31	37	62.7	1280	2	T00365	hypothetical prote
32	37	62.7	2109	2	T33247	hypothetical prote
33	37	62.7	2584	2	T24158	hypothetical prote
34	37	62.7	2606	2	T24157	hypothetical prote
35	37	62.7	3119	2	T18414	protein g377 - mal
36	37	62.7	5138	2	B96695	hypothetical prote
37	36	61.0	182	2	T16423	hypothetical prote
38	36	61.0	206	2	B48441	antigen (C-termina
39	36	61.0	239	1	BVEC	tonB protein - Esc
40	36	61.0	239	2	G85705	hypothetical prote
41	36	61.0	240	2	S13257	tonB protein - Sal
42	36	61.0	243	2	T45505	membrane protein [
43	36	61.0	255	2	S30290	tonB protein - Yer
44	36	61.0	265	2	T00765	hypothetical prote
45	36	61.0	266	2	T44781	tonB protein [impo

ALIGNMENTS

RESULT 1
S35331
apidaecin 22 precursor - honeybee
C;Species: Apis mellifera (honeybee)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C;Accession: S35331
R;Casteels-Josson, K.; Capaci, T.; Casteels, P.; Tempst, P.
EMBO J. 12, 1569-1578, 1993
A;Title: Apidaecin multipeptide precursor structure: a putative mechanism for amplifi
A;Reference number: S35330; MUID:93223697
A;Accession: S35331
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-144 <CAS>
A;Cross-references: EMBL:X72576; NID:g297064; PIDN:CAA51168.1; PID:g297065
C;Superfamily: procyclic acidic repetitive protein

Query Match 71.2%; Score 42; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EAEPEAEP 9
||||||||
Db 35 EAEPEAEP 42

RESULT 2
S35332
apidaecin 73 precursor - honeybee (fragment)
N;Contains: apidaecin Ia
C;Species: Apis mellifera (honeybee)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 03-Nov-2000
C;Accession: S35332; S05383
R;Casteels-Josson, K.; Capaci, T.; Casteels, P.; Tempst, P.
EMBO J. 12, 1569-1578, 1993
A;Title: Apidaecin multipeptide precursor structure: a putative mechanism for amplifi
A;Reference number: S35330; MUID:93223697
A;Accession: S35332
A;Molecule type: mRNA
A;Residues: 1-283 <CAS>
A;Cross-references: EMBL:X72577; NID:g297066; PIDN:CAA51169.1; PID:g4539289
A;Accession: S05383
A;Molecule type: protein
A;Residues: 258-283 <CA3>
C;Superfamily: proline-rich protein
F;266-283/Product: apidaecin Ia #status experimental <MAT>

Query Match 71.2%; Score 42; DB 2; Length 283;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2 EAEPEAEP 9
            ||||||||
Db      62 EAEPEAEP 69
```

RESULT 3
S34224
cyclin B - long-tailed hamster
C;Species: Cricetulus longicaudatus (long-tailed hamster)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C;Accession: S34224
R;Markiewicz, D.A.; Flick, M.B.; Mushel, R.J.; Mckenna, W.G.
submitted to the EMBL Data Library, March 1992
A;Description: New features of mammalian cyclins seen in rat and chinese hamster cyclin
A;Reference number: S34224
A;Accession: S34224
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-429 <MAR>
A;Cross-references: EMBL:X64588; NID:g313764; PIDN:CAA45876.1; PID:g313765
C;Superfamily: cyclin
C;Keywords: cell cycle control

```
  Query Match             67.8%;  Score 40;  DB 2;  Length 429;
  Best Local Similarity   70.0%;  Pred. No. 16;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPIM 11
            | ||| ||:|
Db      98 EPEPEPEPVM 107
```

RESULT 4
T39332
hypothetical protein SPBC11G11.05 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39332
R;Saunders, D.; Harris, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21845
A;Accession: T39332
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-251 <SAU>
A;Cross-references: EMBL:AL132717; PIDN:CAB59807.1; GSPDB:GN00067; SPDB:SPBC11G11.05
A;Experimental source: strain 972h-; cosmid c11G11
C;Genetics:
A;Gene: SPDB:SPBC11G11.05
A;Map position: 2

```
  Query Match             66.1%;  Score 39;  DB 2;  Length 251;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
            |:||||||
Db     185 ESEPEAEP 192
```

RESULT 5
S07633
hematopoietic lineage cell-specific protein HS1 - human
C;Species: Homo sapiens (man)
C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 05-Nov-1999
C;Accession: S07633; A47478; B47478; C47478; D47478; E47478
R;Kitamura, D.; Kaneko, H.; Miyagoe, Y.; Ariyasu, T.; Watanabe, T.
Nucleic Acids Res. 17, 9367-9379, 1989
A;Title: Isolation and characterization of a novel human gene expressed specifically in
A;Reference number: S07633; MUID:90067934
A;Accession: S07633
A;Molecule type: mRNA
A;Residues: 1-486 <KIT>
A;Cross-references: EMBL:X16663; NID:g32054; PIDN:CAA34651.1; PID:g32055
R;Yamanashi, Y.; Okada, M.; Semba, T.; Yamori, T.; Umemori, H.; Tsunasawa, S.; Toyosh
Proc. Natl. Acad. Sci. U.S.A. 90, 3631-3635, 1993
A;Title: Identification of HS1.protein as a major substrate of protein-tyrosine kinas
A;Reference number: A47478; MUID:93234551
A;Accession: A47478
A;Status: preliminary
A;Molecule type: protein
A;Residues: 4-19,'XXX',23-26,'X';79-93,'X',95,'X';134-146;208-223;274-284,'X',286,'X'
A;Experimental source: Daudi, B-lymphoblastoid cells
A;Note: sequence modified after extraction from NCBI backbone
C;Superfamily: SH3 homology
F;435-482/Domain: SH3 homology <SH3>

```
  Query Match             66.1%;  Score 39;  DB 2;  Length 486;
  Best Local Similarity   77.8%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEP 9
            :||||| ||
Db     360 YEAEPEPEP 368
```

RESULT 6
T23694
hypothetical protein M03C11.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23694
R;McMurray, A.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19783
A;Accession: T23694
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-848 <WIL>
A;Cross-references: EMBL:Z49128; PIDN:CAA88959.1; GSPDB:GN00021; CESP:M03C11.2
A;Experimental source: clone M03C11
C;Genetics:
A;Gene: CESP:M03C11.2
A;Map position: 3
A;Introns: 113/2; 147/3; 185/3; 379/1; 482/3; 553/2; 688/3; 762/3

```
  Query Match             66.1%;  Score 39;  DB 2;  Length 848;
  Best Local Similarity   77.8%;  Pred. No. 48;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPI 10
            | ||||||:
Db     506 EPEPEAEPL 514
```

RESULT 7
T22028
hypothetical protein F40F11.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22028
R;Dobson, R.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z19504
A;Accession: T22028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1593 <WIL>
A;Cross-references: EMBL:Z73426; NID:e1343215; PIDN:CAA97793.1; GSPDB:GN00022; CESP:F
A;Experimental source: clone F40F11
C;Genetics:

```
A;Gene: CESP:F40F11.2
A;Map position: 4
A;Introns: 22/3; 96/3; 129/3; 206/1; 248/3; 710/1; 811/1; 1517/3; 1552/1


  Query Match             66.1%;  Score 39;  DB 2;  Length 1593;
  Best Local Similarity   87.5%;  Pred. No. 93;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
           |:||||||
Db     851 ESEPEAEP 858


RESULT   8
C72724
hypothetical protein APE0332 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: C72724
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: C72724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-150 <KAW>
A;Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79287.1; PID:d1043073; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0332
C;Superfamily: Aeropyrum pernix hypothetical protein APE0332


  Query Match             64.4%;  Score 38;  DB 2;  Length 150;
  Best Local Similarity   66.7%;  Pred. No. 12;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 AEPEAEPIM 11
           |||| ||::
Db      45 AEPEVEPVL 53


RESULT   9
S35330
apidaecin 14 precursor - honeybee
N;Contains: apidaecin II
C;Species: Apis mellifera (honeybee)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C;Accession: S35330; S06676
R;Casteels-Josson, K.; Capaci, T.; Casteels, P.; Tempst, P.
EMBO J. 12, 1569-1578, 1993
A;Title: Apidaecin multipeptide precursor structure: a putative mechanism for amplificat
A;Reference number: S35330; MUID:93223697
A;Accession: S35330
A;Molecule type: mRNA
A;Residues: 1-168 <CAS>
A;Cross-references: EMBL:X72575; NID:g297062; PIDN:CAA51167.1; PID:g297063
R;Casteels, P.; Ampe, C.; Jacobs, F.; Vaeck, M.; Tempst, P.
EMBO J. 8, 2387-2391, 1989
A;Title: Apidaecins: antibacterial peptides from honeybees.
A;Reference number: S05383; MUID:90005446
A;Accession: S06676
A;Molecule type: protein
A;Residues: 43-60 <CA2>
C;Superfamily: procyclic acidic repetitive protein
F;43-60/Product: apidaecin II #status experimental <MAT>


  Query Match             64.4%;  Score 38;  DB 2;  Length 168;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
           ||:|||||
Db      35 EAKPEAEP 42


RESULT  10
S33755
granzyme-like protein I (EC 3.4.21.-) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 22-Jun-1999
C;Accession: S33755
R;Amerik, A.Y.; Yarovoi, S.V.; Grigorenko, V.G.; Antonov, V.K.
FEBS Lett. 324, 226-230, 1993
A;Title: Identification, sequence analysis, and characterization of cDNA clones encod
A;Reference number: S33755; MUID:93285341
A;Accession: S33755
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-248 <AME>
A;Cross-references: EMBL:X66693; NID:g296177; PIDN:CAA47235.1; PID:g296178
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;21-241/Domain: trypsin homology <TRY>
F;65,109,204/Active site: His, Asp, Ser #status predicted


  Query Match             64.4%;  Score 38;  DB 2;  Length 248;
  Best Local Similarity   54.5%;  Pred. No. 20;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      25 HEADPHSRPYM 35


RESULT  11
PRHU3
proteinase 3 (EC 3.4.21.-) precursor [validated] - human
N;Alternate names: AGP7; C-ANCA antigen; neutrophil proteinase 4; p29; Wegener's gran
N;Contains: myeloblastin
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 08-Dec-2000
C;Accession: A45080; B46268; A43983; JH0331; A33751; S11091; A61176; A60148; A43982;
R;Sturrock, A.B.; Franklin, K.F.; Rao, G.; Marshall, B.C.; Rebentisch, M.B.; Lemons,
J. Biol. Chem. 267, 21193-21199, 1992
A;Title: Structure, chromosomal assignment, and expression of the gene for proteinase
A;Reference number: A45080; MUID:93016043
A;Accession: A45080
A;Molecule type: DNA
A;Residues: 1-254,'P' <STU>
A;Cross-references: GB:M97911
A;Note: sequence extracted from NCBI backbone (NCBIP:116205)
R;Zimmer, M.; Medcalf, R.L.; Fink, T.M.; Mattmann, C.; Lichter, P.; Jenne, D.E.
Proc. Natl. Acad. Sci. U.S.A. 89, 8215-8219, 1992
A;Title: Three human elastase-like genes coordinately expressed in the myelomonocyte
A;Reference number: A46268; MUID:92390417
A;Accession: B46268
A;Molecule type: DNA
A;Residues: 1-118,'V',120-134,'AT',137-256 <ZIM>
A;Note: sequence extracted from NCBI backbone (NCBIN:112898, NCBIN:112900, NCBIN:1129
R;Labbaye, C.; Musette, P.; Cayre, Y.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 9253-9256, 1991
A;Title: Wegener autoantigen and myeloblastin are encoded by a single mRNA.
A;Reference number: A43983; MUID:92021028
A;Accession: A43983
A;Molecule type: mRNA
A;Residues: 1-69,'P',71-256 <LA2>
A;Cross-references: GB:M75154; NID:g187398; PIDN:AAA59558.1; PID:g187399
R;Campanelli, D.; Melchior, M.; Fu, Y.; Nakata, M.; Shuman, H.; Nathan, C.; Gabay, J.
J. Exp. Med. 172, 1709-1715, 1990
A;Title: Cloning of cDNA for proteinase 3: a serine protease, antibiotic, and autoant
```

A;Reference number: JH0331; MUID:91079774
A;Accession: JH0331
A;Molecule type: mRNA
A;Residues: 'R',3-118,'V',120-134,'AT',137-254,'P' <CAM>
A;Cross-references: GB:X55668; NID:g35687; PIDN:CAA39203.1; PID:g1335280
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Bories, D.; Raynal, M.C.; Solomon, D.H.; Darzynkiewicz, Z.; Cayre, Y.E.
Cell 59, 959-968, 1989
A;Title: Down-regulation of a serine protease, myeloblastin, causes growth arrest and di
A;Reference number: A33751; MUID:90090622
A;Accession: A33751
A;Molecule type: mRNA
A;Residues: 42-256 <BOR>
A;Cross-references: GB:M29142; NID:g188983; PIDN:AAA36342.1; PID:g188984
A;Note: the authors translated the codon GGG for residue 49 as Glu, GGC for residue 52 a
R;Jenne, D.E.; Tschopp, J.; Luedemann, J.; Utecht, B.; Gross, W.L.
Nature 346, 520, 1990
A;Title: Wegener's autoantigen decoded.
A;Reference number: S11091; MUID:90332035
A;Accession: S11091
A;Molecule type: mRNA
A;Residues: 20-56 <JEN>
R;Musette, P.; Labbaye, C.; Dorner, M.H.; Cayre, Y.E.; Casanova, J.L.; Kourilsky, P.
Blood 77, 1398-1399, 1991
A;Title: Wegener's autoantigen and leukemia.
A;Reference number: A61176; MUID:91159650
A;Accession: A61176
A;Molecule type: mRNA
A;Residues: 1-42 <MUS>
A;Cross-references: EMBL:X56606; NID:g35189; PIDN:CAA39943.1; PID:g35190
R;Goldschmeding, R.; Dolman, K.M.; Van Den Ende, M.E.; Van Der Meer-Gerritsen, C.H.; Son
APMIS 19(Suppl.), 26-27, 1990
A;Title: The relation of 29 kD C-ANCA antigen to proteinase 3.
A;Reference number: A60148; MUID:91136884
A;Accession: A60148
A;Molecule type: protein
A;Residues: 28-48 <GOL>
R;Rao, N.V.; Wehner, N.G.; Marshall, B.C.; Gray, W.R.; Gray, B.H.; Hoidal, J.R.
J. Biol. Chem. 266, 9540-9548, 1991
A;Title: Characterization of proteinase-3 (PR-3), a neutrophil serine proteinase. Struct
A;Reference number: A43982; MUID:91236723
A;Accession: A43982
A;Molecule type: protein
A;Residues: 28-61,'X',63,'D',65-67;228-244 <RAO>
R;Wilde, C.G.; Snable, J.L.; Griffith, J.E.; Scott, R.W.
J. Biol. Chem. 265, 2038-2041, 1990
A;Title: Characterization of two azurophil granule proteases with active-site homology t
A;Reference number: A43981; MUID:90130450
A;Accession: A43981
A;Molecule type: protein
A;Residues: 28-45,'E',47;196-208,'X',210-215,'X',217-219 <WIL>
R;Gabay, J.E.; Scott, R.W.; Campanelli, D.; Griffith, J.; Wilde, C.; Marra, M.N.; Seeger
Proc. Natl. Acad. Sci. U.S.A. 86, 5610-5614, 1989
A;Title: Antibiotic proteins of human polymorphonuclear leukocytes.
A;Reference number: A33913; MUID:89315847
A;Accession: C33913
A;Molecule type: protein
A;Residues: 28-45,'E',47 <GAB>
R;Niles, J.L.; McCluskey, R.T.; Ahmad, M.F.; Arnaout, M.A.
Blood 74, 1888-1893, 1989
A;Title: Wegener's granulomatosis autoantigen is a novel neutrophil serine proteinase.
A;Reference number: A60481; MUID:90028708
A;Accession: A60481
A;Molecule type: protein
A;Residues: 28-38,'X',40-47 <NI2>
R;Ohlsson, K.; Linder, C.; Rosengren, M.
Biol. Chem. Hoppe-Seyler 371, 549-555, 1990
A;Title: Monoclonal antibodies specific for neutrophil proteinase 4. Production and use
A;Reference number: S10605; MUID:91025622
A;Accession: S10605
A;Molecule type: protein
A;Residues: 28-52 <OHL>
R;Luedemann, J.; Utecht, B.; Gross, W.L.
J. Exp. Med. 171, 357-362, 1990
A;Title: Anti-neutrophil cytoplasm antibodies in Wegener's granulomatosis recognize a
A;Reference number: PL0230; MUID:90111630
A;Accession: PL0230
A;Molecule type: protein
A;Residues: 28-37,'I',39-40,'I',41-43 <LUE>
C;Comment: This polymorphonuclear leukocyte serine protease from azurophilic granules
C;Genetics:
A;Gene: GDB:PRTN3
A;Cross-references: GDB:126876; OMIM:177020
A;Map position: 19p13.3-19p13.3
A;Introns: 21/1; 76/2; 123/3; 200/3
C;Superfamily: trypsin; trypsin homology
C;Keywords: glycoprotein; hydrolase; polymorphonuclear leukocyte; serine proteinase
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-27/Domain: propeptide #status predicted <PRO>
F;28-256/Product: proteinase 3 #status experimental <MAT>
F;28-243/Domain: trypsin homology <TRY>
F;56-72,152-209,182-188,199-224/Disulfide bonds: #status predicted
F;71,118,203/Active site: His, Asp, Ser #status predicted
F;129,174/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 64.4%; Score 38; DB 1; Length 256;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      32 HEAQPHSRPYM 42
```

RESULT 12
T51059
hypothetical protein B12F1.110 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51059
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51059
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-288 <SCH>
A;Cross-references: EMBL:AL390091; GSPDB:GN00116; NCSP:B12F1.110
A;Experimental source: BAC clone B12F1; strain OR74A
C;Genetics:
A;Gene: NCSP:B12F1.110
A;Map position: 6
A;Introns: 154/1

Query Match 64.4%; Score 38; DB 2; Length 288;
Best Local Similarity 54.5%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 HEAEPEAEPIM 11
           ||  ||: ||:
Db      58 HEPHPESHPIL 68
```

RESULT 13
S31848
heat shock protein HSP30 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YCR021c
C;Species: Saccharomyces cerevisiae
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 29-Oct-1999
C;Accession: S31848; S19432; S30781
R;Regnacq, M.; Boucherie, H.
Curr. Genet. 23, 435-442, 1993
A;Title: Isolation and sequence of HSP30, a yeast heat-shock gene coding for a hydrop
A;Reference number: S31848; MUID:93306747

A;Accession: S31848
A;Molecule type: DNA
A;Residues: 1-263 <REG1>
A;Cross-references: EMBL:M93123
R;Feldmann, H.; Mannhaupt, G.; Vetter, I.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19429
A;Accession: S19432
A;Molecule type: DNA
A;Residues: 1-190,'A',192-332 <FEL>
A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42313.1; PID:e264485; PID:g190716
R;Regnacq, M.; Boucherie, H.
submitted to the EMBL Data Library, January 1993
A;Reference number: S30781
A;Accession: S30781
A;Molecule type: DNA
A;Residues: 1-152,'GY',164-165,167,'A',169,'NSNRGYL',170-240,'VFNQT',270,278-280,'FMVFWI
A;Cross-references: EMBL:M93123
A;Note: the difference at the carboxyl end is due to a frameshift error
C;Genetics:
A;Gene: SGD:HSP30
A;Cross-references: SGD:S0000615; MIPS:YCR021c
A;Map position: 3R
C;Keywords: membrane protein

Query Match 64.4%; Score 38; DB 2; Length 332;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HEAEPEAE 8
|| |||||
Db 318 HEPEPEAE 325

RESULT 14
S69371
duodenase - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 21-Nov-1998
C;Accession: S69371
R;Zamolodchikova, T.S.; Vorotyntseva, T.I.; Antonov, V.K.
Eur. J. Biochem. 227, 866-872, 1995
A;Title: Duodenase, a new serine protease of unusual specificity from bovine duodenal mu
A;Reference number: S69371; MUID:95172075
A;Accession: S69371
A;Molecule type: protein
A;Residues: 1-21 <ZAM>
C;Superfamily: trypsin; trypsin homology

Query Match 62.7%; Score 37; DB 2; Length 21;
Best Local Similarity 54.5%; Pred. No. 2.3;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 HEAEPEAEPIM 11
|||:| : | |
Db 5 HEAKPHSRPYM 15

RESULT 15
T36257
hypothetical protein SCE68.07c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36257
R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A;Reference number: Z21576
A;Accession: T36257
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-125 <MUR>
A;Cross-references: EMBL:AL079345; PIDN:CAB45343.1; GSPDB:GN00070; SCOEDB:SCE68.07c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE68.07c

Query Match 62.7%; Score 37; DB 2; Length 125;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EAEPEAEP 9
||||| ||
Db 5 EAEPEPEP 12

Search completed: January 4, 2002, 08:41:31
Job time: 169 sec

```
OM of: US-09-444-281-27 to: Issued_Patents_NA:*   out_format : pfs

Date: Jan 4, 2002 10:57 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084144_16216/app_query.fasta_1.210
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09444281_@CGN1_1_70 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-27
Query length: 11
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 135.410000


score_list:
Sequence            Strd Orig   ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2 +   42.00   53.44  2.1e+05   4403765
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1 +   42.00   53.43  2.1e+05   4411529
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-479-537A-1 -   39.00  100.20  624.61    6192  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-116-1 -    39.00  100.20  624.61    6192  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-479-537A-4 -   39.00   99.85  653.38    6449  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-116-4 -   39.00   99.85  653.38    6449  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-394-600B-12 -   38.00  134.11  8.07       83  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02513-12 -   38.00  134.11  8.07      83
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-394-600B-3 +   38.00  114.87  95.13     771  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-394-600B-22 +   38.00  114.87  95.13     771  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-230-428B-3 +   38.00  114.87  95.13     771  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02513-3 +   38.00  114.87  95.13     771
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02513-22 +   38.00  114.87  95.13     771
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-549A-645 +   37.00  130.10  13.49      90  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-549A-646 -   37.00  130.10  13.49      90  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-592-126-120 +   37.00  119.04  55.71     324  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-851-974-2 +   37.00  110.38  169.25    884  !  S
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-213-390-2 +   37.00  110.38  169.25    884  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-434-288-7 -   37.00  104.83  344.79    1681  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-178-252-3 -   37.00  104.17  375.35    1815  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-178-252-7 -   37.00  104.17  375.35    1815  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-178-252-8 -   37.00  104.17  375.35    1815  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-996-139-14 +   37.00  103.87  389.80    1878  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-995-659-14 +   37.00  103.87  389.80    1878  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-215-649A-14 +   37.00  103.87  389.80    1878  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-899-437-22 -   37.00  102.95  439.04    2091  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-126-121-22 -   37.00  102.95  439.04    2091  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-899-437-5 -   37.00  101.40  535.53    2502  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-126-121-5 -   37.00  101.40  535.53    2502  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-899-437-1 -   37.00  101.27  544.07    2538  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-126-121-1 -   37.00  101.27  544.07    2538  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-247-946A-2 -   37.00   98.88  739.81    3350  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-06420-2 -   37.00   98.88  739.81    3350
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-549A-639 -   37.00   98.76  751.07    3396
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-178-252-1 -   37.00   98.64  762.83    3444  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-549A-721 -   37.00   98.62  764.54    3451
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-179-189 +   36.00  113.43  114.44     423  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-651-940-1 +   36.00  102.49  465.44    1502  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-286-904-23 +   36.00   99.28  703.04    2180  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-120-663-1 +   36.00   91.26  2.0e+03   5517  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-658-322-4 -   35.00  126.42  21.63      64  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-923-856-2 +   35.00  101.29  542.82    1176
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-216-294-2 +   35.00  101.29  542.82    1176
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-425-299A-1 +   35.00  100.41  607.57    1302
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-187-785-2 +   35.00   99.47  685.52    1452

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis .
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
           Quality:   42.00            Length:      9
             Ratio:  4.667              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-27 x US-09-103-840A-2   ..

 Align seg 1/1  to: US-09-103-840A-2  from: 1  to: 4403765

       3 AlaGluProGluAlaGluProIleMet 11
         |||||||||||||||||||||||:::
  260647 GCTGAACCGGAGGCAGAGCCGATCGTC 260673

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
alignment_scores:
            Quality:   42.00            Length:      9
              Ratio:  4.667              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-27 x US-09-103-840A-1    ..

 Align seg 1/1  to: US-09-103-840A-1  from: 1  to: 4411529

       3 AlaGluProGluAlaGluProIleMet 11
         ||||||||||||||||||||||||:::
  260536 GCTGAACCGGAGGCAGAGCCGATCGTC 260562

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-479-537A-1

seq_documentation_block:
; Sequence 1, Application US/08479537A
; Patent No. 5861381
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,537A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90/13101
;      FILING DATE:  23-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6192 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  58..120
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  439..5239
;      OTHER INFORMATION:  /note= "The nucleotides spanning
;      OTHER INFORMATION:  439-5239 constitute a repeated region wherein the repeat i
;      OTHER INFORMATION:  nucleotides and encodes 20 amino acids, 17 of which are fi
;      OTHER INFORMATION:  The number of such repeats varies from 1 to 80."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  121..6166
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  457
;      OTHER INFORMATION:  /note= "Nucleotide 457 is X1 = NNN
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  487
;      OTHER INFORMATION:  /note= "Nucleotide 487 is Y = NNN
;      OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;      OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  496
;      OTHER INFORMATION:  /note= "Nucleotide 496 is X2 = NNN
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-1


alignment_scores:
            Quality:   39.00            Length:      8
              Ratio:  4.875              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-444-281-27 x US-08-479-537A-1/rev    ..

 Align seg 1/1  to reverse of: US-08-479-537A-1  from: 1  to: 6192

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||:::|||||||||
    5306 GAAGCTGAGCCTGATGCAGAGCCT 5283

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-116-1

seq_documentation_block:
; Sequence 1, Application US/09083116
; Patent No. 6203795
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/083,116
;      FILING DATE:
;      CLASSIFICATION:
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/479,537
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/FR91/00835
;       FILING DATE:  23-OCT-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/039,320
;       FILING DATE:  04-APR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/403,576
;       FILING DATE:  14-MAR-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
;       REFERENCE/DOCKET NUMBER:  017753-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6192 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  58..120
;     FEATURE:
;       NAME/KEY:  repeat_region
;       LOCATION:  439..5239
;       OTHER INFORMATION:  /note= "The nucleotides spanning
;       OTHER INFORMATION:  439-5239 constitute a repeated region wherein the repeat is 6
;       OTHER INFORMATION:  nucleotides and encodes 20 amino acids, 17 of which are fixed
;       OTHER INFORMATION:  The number of such repeats varies from 1 to 80."
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  121..6166
;     FEATURE:
;       NAME/KEY:  repeat_region
;       LOCATION:  457
;       OTHER INFORMATION:  /note= "Nucleotide 457 is X1 = NNN
;       OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
;       OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;     FEATURE:
;       NAME/KEY:  repeat_region
;       LOCATION:  487
;       OTHER INFORMATION:  /note= "Nucleotide 487 is Y = NNN
;       OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
;       OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;     FEATURE:
;       NAME/KEY:  repeat_region
;       LOCATION:  496
;       OTHER INFORMATION:  /note= "Nucleotide 496 is X2 = NNN
;       OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
;       OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-09-083-116-1


alignment_scores:
            Quality:   39.00              Length:      8
              Ratio:  4.875                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-444-281-27 x US-09-083-116-1/rev

 Align seg 1/1  to reverse of: US-09-083-116-1  from: 1  to: 6192

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||:::|||||||||
    5306 GAAGCTGAGCCTGATGCAGAGCCT 5283
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-479-537A-4

seq_documentation_block:
; Sequence 4, Application US/08479537A
; Patent No. 5861381
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,537A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90/13101
;      FILING DATE:  23-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6449 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  58..120
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  439..5239
;      OTHER INFORMATION:  /note= "The nucleotides spanning
;      OTHER INFORMATION:  439-5239 constitute a repeated region wherein the repeat i
;      OTHER INFORMATION:  nucleotides and encodes 20 amino acids, 17 of which are fi
;      OTHER INFORMATION:  The number of such repeats varies from 1 to 80."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  121..5661
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  457
;      OTHER INFORMATION:  /note= "Nucleotide 457 is X1 = NNN
```

```
;       OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
;       OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;     FEATURE:
;       NAME/KEY:  repeat_region
;       LOCATION:  487
;       OTHER INFORMATION:  /note= "Nucleotide 487 is Y = NNN
;       OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
;       OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;     FEATURE:
;       NAME/KEY:  repeat_region
;       LOCATION:  496
;       OTHER INFORMATION:  /note= "Nucleotide 496 is X2 = NNN
;       OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
;       OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-08-479-537A-4


alignment_scores:
            Quality:   39.00              Length:      8
              Ratio:  4.875                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-444-281-27 x US-08-479-537A-4/rev    ..

 Align seg 1/1  to reverse of: US-08-479-537A-4  from: 1  to: 6449

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||:::|||||||||
    5306 GAAGCTGAGCCTGATGCAGAGCCT 5283

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-116-4

seq_documentation_block:
; Sequence 4, Application US/09083116
; Patent No. 6203795
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/083,116
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/479,537
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6449 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  58..120
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  439..5239
;      OTHER INFORMATION:  /note= "The nucleotides spanning
;      OTHER INFORMATION:  439-5239 constitute a repeated region wherein the repeat i
;      OTHER INFORMATION:  nucleotides and encodes 20 amino acids, 17 of which are fi
;      OTHER INFORMATION:  The number of such repeats varies from 1 to 80."
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  121..5661
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  457
;      OTHER INFORMATION:  /note= "Nucleotide 457 is X1 = NNN
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  487
;      OTHER INFORMATION:  /note= "Nucleotide 487 is Y = NNN
;      OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;      OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  496
;      OTHER INFORMATION:  /note= "Nucleotide 496 is X2 = NNN
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
US-09-083-116-4


alignment_scores:
            Quality:   39.00              Length:      8
              Ratio:  4.875                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-444-281-27 x US-09-083-116-4/rev    ..

 Align seg 1/1  to reverse of: US-09-083-116-4  from: 1  to: 6449

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||:::|||||||||
    5306 GAAGCTGAGCCTGATGCAGAGCCT 5283

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-394-600B-12

seq_documentation_block:
; Sequence 12, Application US/08394600B
; Patent No. 5843693
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
```

```
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McAndrews, Held & Malloy, Ltd.
;      STREET:  500 West Madison Street; 34th Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60661
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/394,600B
;      FILING DATE:  02/27/95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donald J. Pochopien
;      REGISTRATION NUMBER:  32,167
;      REFERENCE/DOCKET NUMBER:  820.005/11850US05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/707-8889
;      TELEFAX:  312/707-9155
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  83 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-394-600B-12


alignment_scores:
            Quality:   38.00             Length:     11
              Ratio:  3.800               Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x US-08-394-600B-12/rev   ..

 Align seg 1/1  to reverse of: US-08-394-600B-12  from: 1  to: 83

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::|||::::::::::|||   |||
      65 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 33

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02513-12

seq_documentation_block:
; Sequence 12, Application PC/TUS9502513
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02513
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams Jr., Joseph A.
;      REGISTRATION NUMBER:  38,659
;      REFERENCE/DOCKET NUMBER:  27527/32404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  83 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-02513-12


alignment_scores:
            Quality:   38.00             Length:     11
              Ratio:  3.800               Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x PCT-US95-02513-12/rev   ..

 Align seg 1/1  to reverse of: PCT-US95-02513-12  from: 1  to: 83

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::|||::::::::::|||   |||
      65 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 33

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-394-600B-3

seq_documentation_block:
; Sequence 3, Application US/08394600B
; Patent No. 5843693
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McAndrews, Held & Malloy, Ltd.
;      STREET:  500 West Madison Street; 34th Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60661
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/394,600B
;      FILING DATE:  02/27/95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donald J. Pochopien
;      REGISTRATION NUMBER:  32,167
;      REFERENCE/DOCKET NUMBER:  820.005/11850US05
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/707-8889
;       TELEFAX:  312/707-9155
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  771 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..771
US-08-394-600B-3


alignment_scores:
            Quality:   38.00              Length:     11
              Ratio:  3.800                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x US-08-394-600B-3    ..

 Align seg 1/1  to: US-08-394-600B-3  from: 1  to: 771

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::|||::::::::::|||   |||
      94 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 126

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-394-600B-22

seq_documentation_block:
; Sequence 22, Application US/08394600B
; Patent No. 5843693
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McAndrews, Held & Malloy, Ltd.
;      STREET:  500 West Madison Street; 34th Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60661
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/394,600B
;      FILING DATE:  02/27/95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donald J. Pochopien
;      REGISTRATION NUMBER:  32,167
;      REFERENCE/DOCKET NUMBER:  820.005/11850US05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/707-8889
;      TELEFAX:  312/707-9155
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..771
US-08-394-600B-22


alignment_scores:
            Quality:   38.00              Length:     11
              Ratio:  3.800                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x US-08-394-600B-22    ..

 Align seg 1/1  to: US-08-394-600B-22  from: 1  to: 771

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::|||::::::::::|||   |||
      94 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 126

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-230-428B-3

seq_documentation_block:
; Sequence 3, Application US/08230428B
; Patent No. 5998378
;  GENERAL INFORMATION:
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    TITLE OF INVENTION:  Compositions For The Inhibition Of TNF
;    TITLE OF INVENTION:  Hormone Formation And Uses Thereof (As Amended)
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION  Intellectual Property - R440
;      STREET:  4560 Horton Street, P.O. Box 8097
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,428B
;      FILING DATE:  19-APR-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/905,546
;      FILING DATE:  25-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/395,253
;      FILING DATE:  16-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savereide, Paul B.
;      REGISTRATION NUMBER:  36,914
;      REFERENCE/DOCKET NUMBER:  0820.004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2718
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..768
US-08-230-428B-3


alignment_scores:
            Quality:   38.00              Length:     11
              Ratio:  3.800                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x US-08-230-428B-3   ..

 Align seg 1/1  to: US-08-230-428B-3  from: 1  to: 771

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::|||::::::::::|||   |||
      94 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 126

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02513-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9502513
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02513
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams Jr., Joseph A.
;      REGISTRATION NUMBER:  38,659
;      REFERENCE/DOCKET NUMBER:  27527/32404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..771
PCT-US95-02513-3


alignment_scores:
            Quality:   38.00              Length:     11
              Ratio:  3.800                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x PCT-US95-02513-3   ..

 Align seg 1/1  to: PCT-US95-02513-3  from: 1  to: 771

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::|||::::::::::|||   |||
      94 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 126

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02513-22

seq_documentation_block:
; Sequence 22, Application PC/TUS9502513
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02513
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams Jr., Joseph A.
;      REGISTRATION NUMBER:  38,659
;      REFERENCE/DOCKET NUMBER:  27527/32404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..771
PCT-US95-02513-22


alignment_scores:
            Quality:   38.00              Length:     11
              Ratio:  3.800                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 54.545

alignment_block:
 US-09-444-281-27 x PCT-US95-02513-22   ..

 Align seg 1/1  to: PCT-US95-02513-22  from: 1  to: 771
```

```
      1 HisGluAlaGluProGluAlaGluProIleMet 11
        ||||||||||:::|||::::::::::|||. |||
     94 CACGAGGCGCAGCCACACTCCCGGCCCTACATG 126

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-549A-645

seq_documentation_block:
; Sequence 645, Application US/08974549A
; Patent No. 6166178
;  GENERAL INFORMATION:
;    APPLICANT:  Cech, Thomas R.
;    APPLICANT:  Lingner, Joachim
;    APPLICANT:  Nakamura, Toru
;    APPLICANT:  Chapman, Karen B.
;    APPLICANT:  Morin, Gregg B.
;    APPLICANT:  Harley, Calvin B.
;    APPLICANT:  Andrews, William H.
;    TITLE OF INVENTION:  Human Telomerase Catalytic Subunit
;    NUMBER OF SEQUENCES:  727
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/974,549A
;      FILING DATE:  19-NOV-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/724,643
;      FILING DATE:  01-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/844,419
;      FILING DATE:  18-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/846,017
;      FILING DATE:  25-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/851,843
;      FILING DATE:  06-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/854,050
;      FILING DATE:  09-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/911,312
;      FILING DATE:  14-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/912,951
;      FILING DATE:  14-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/915,503
;      FILING DATE:  14-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US97/17618
;      FILING DATE:  01-OCT-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US97/17885
;      FILING DATE:  01-OCT-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Apple, Randolph Ted
;      REGISTRATION NUMBER:  36,429
;      REFERENCE/DOCKET NUMBER:  015389-002610US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  645:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  90 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..90
;      OTHER INFORMATION:  /note= "oligonucleotide 2B"
US-08-974-549A-645


alignment_scores:
            Quality:   37.00              Length:       9
              Ratio:  4.111                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
 US-09-444-281-27 x US-08-974-549A-645   ..

 Align seg 1/1  to: US-08-974-549A-645  from: 1  to: 90

       1 HisGluAlaGluProGluAlaGluPro 9
         |||:::|||||||||:::|||:::|||
      37 CACCACGCTGAACCAGACGCCAACCCT 63

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-549A-646

seq_documentation_block:
; Sequence 646, Application US/08974549A
; Patent No. 6166178
;  GENERAL INFORMATION:
;    APPLICANT:  Cech, Thomas R.
;    APPLICANT:  Lingner, Joachim
;    APPLICANT:  Nakamura, Toru
;    APPLICANT:  Chapman, Karen B.
;    APPLICANT:  Morin, Gregg B.
;    APPLICANT:  Harley, Calvin B.
;    APPLICANT:  Andrews, William H.
;    TITLE OF INVENTION:  Human Telomerase Catalytic Subunit
;    NUMBER OF SEQUENCES:  727
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/974,549A
;      FILING DATE:  19-NOV-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/724,643
;      FILING DATE:  01-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/844,419
;      FILING DATE:  18-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/846,017
;      FILING DATE:  25-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/851,843
;      FILING DATE:  06-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/854,050
;      FILING DATE:  09-MAY-1997
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/911,312
;       FILING DATE:  14-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/912,951
;       FILING DATE:  14-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/915,503
;       FILING DATE:  14-AUG-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US97/17618
;       FILING DATE:  01-OCT-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US97/17885
;       FILING DATE:  01-OCT-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Apple, Randolph Ted
;       REGISTRATION NUMBER:  36,429
;       REFERENCE/DOCKET NUMBER:  015389-002610US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 576-0200
;       TELEFAX:  (415) 576-0300
;   INFORMATION FOR SEQ ID NO:  646:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  90 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     FEATURE:
;       NAME/KEY:  -
;       LOCATION:  1..90
;       OTHER INFORMATION:  /note= "oligonucleotide 2T"
US-08-974-549A-646


alignment_scores:
            Quality:   37.00          Length:      9
              Ratio:  4.111            Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 66.667

alignment_block:
 US-09-444-281-27 x US-08-974-549A-646/rev   ..

 Align seg 1/1  to reverse of: US-08-974-549A-646  from: 1  to: 90

       1 HisGluAlaGluProGluAlaGluPro 9
         |||:::||||||||:::|||:::|||
      64 CACCACGCTGAACCAGACGCGAACCCT 38
```

OM of: US-09-444-281-27 to: GenEmbl:* out_format : pfs

Date: Jan 4, 2002 10:54 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084143_16195/app_query.fasta_1.210
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09444281_@CGN1_1_5145 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-27
Query length: 11
Database: GenEmbl:*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1728.450000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_ph:AF109874	-	46.00	111.19	645.10	38347	! AF109874 Bacteriophage Tuc200
gb_pr:AL391821	-	44.00	95.22	5.0e+03	158408	! AL391821 Human DNA sequence
gb_htg:AC021098	+	44.00	94.76	5.3e+03	169671	! AC021098 Homo sapiens chromo
gb_ro:RRPXFPRT	+	43.00	126.67	88.61	966	! Y09049 R.rattus mRNA for PxF pr
gb_ba:AF127222	-	43.00	120.21	202.80	2501	! AF127222 Pseudomonas fluoresce
gb_htg:AC015790	-	43.00	96.71	4.1e+03	79795	! AC015790 Homo sapiens clone R
gb_pr:AL138709	-	43.00	95.39	4.9e+03	96874	! AL138709 Human DNA sequence f
gb_htg:AC019290	-	43.00	91.31	8.3e+03	176712	! AC019290 Homo sapiens clone
gb_htg:AC016270	+	43.00	90.86	8.8e+03	189022	! AC016270 Homo sapiens clone
gb_in:AMAPID22	+	42.00	126.65	88.78	607	! X72576 A.mellifera Apid22 mRNA.
gb_pl:AF034945	-	42.00	124.86	111.67	790	! AF034945 Zea mays glycine-rich
gb_ro:MMY12380	+	42.00	123.87	126.90	915	! Y12380 M.musculus ITGA7 gene, p
gb_in:AMAPID73	+	42.00	123.20	138.29	1010	! X72577 Apis mellifera Apid73 m
gb_in:DDU66367	-	42.00	121.03	182.60	1390	! U66367 Dictyostelium discoideu
gb_in:DDU72746	-	42.00	120.57	193.53	1486	! U72746 Dictyostelium discoideu
gb_pr:HSSTPKEMK	-	42.00	115.93	351.11	2946	! X97630 H.sapiens mRNA for seri
gb_ro:AB01015S09	-	42.00	113.21	497.53	4397	! AB012588 Mus musculus Munc18-1
gb_htg:AC017989	+	42.00	106.60	1.2e+03	11643	! AC017989 Drosophila melanogas
gb_ba:AE006932	+	42.00	104.02	1.6e+03	17042	! AE006932 Mycobacterium tuberc
gb_pr:U73647	+	42.00	98.46	3.3e+03	38639	! U73647 Human Chromosome 11 Co
gb_ba:MTCY8D5	+	42.00	98.27	3.4e+03	39730	! Z92669 Mycobacterium tubercul
gb_ba:MSGY140	-	42.00	98.09	3.5e+03	40806	! AD000006 Mycobacterium tuberc
gb_in:CELT23E7	+	42.00	97.33	3.8e+03	45635	! AF026205 Caenorhabditis elega
gb_htg:AC087785	-	42.00	94.99	5.1e+03	64427	! AC087785 Homo sapiens chromos
gb_htg:AC090264	-	42.00	94.71	5.3e+03	67203	! AC090264 Homo sapiens chromos
gb_htg:AC010702	-	42.00	94.58	5.4e+03	68457	! AC010702 Drosophila melanogas
gb_htg:AC010571	-	42.00	91.67	7.9e+03	105160	! AC010571 Drosophila melanoga
gb_htg:AC017581	-	42.00	91.01	8.6e+03	115884	! AC017581 Drosophila melanoga
gb_htg:AC091284	-	42.00	90.89	8.7e+03	117993	! AC091284 Mus musculus clone
gb_htg:AL450398	+	42.00	89.19	1.1e+04	151608	! AL450398 Homo sapiens chromo
gb_htg:AP003780	-	42.00	89.14	1.1e+04	152607	! AP003780 Homo sapiens chromo
gb_htg:AC013253	+	42.00	88.94	1.1e+04	157108	! AC013253 Homo sapiens chromo
gb_htg:AC092730	-	42.00	88.93	1.1e+04	157505	! AC092730 Felis catus clone R
gb_pr:AC005242	-	42.00	88.81	1.1e+04	160262	! AC005242 Homo sapiens chromo
gb_htg:AL358034	-	42.00	88.76	1.1e+04	161388	! AL358034 Homo sapiens chromo
gb_in:AC008228	+	42.00	88.70	1.2e+04	162951	! AC008228 Drosophila melanoga
gb_htg:AC022490	+	42.00	88.37	1.2e+04	170892	! AC022490 Homo sapiens chromo
gb_htg:AC010195	-	42.00	88.22	1.2e+04	174852	! AC010195 Homo sapiens clone
gb_htg:AC023920	+	42.00	87.57	1.3e+04	192453	! AC023920 Homo sapiens clone
gb_in:AE003594	+	42.00	84.43	2.0e+04	305516	! AE003594 Drosophila melanoga
gb_in:AE003587	+	42.00	84.35	2.0e+04	309156	! AE003587 Drosophila melanoga
gb_ba:AP002996	-	42.00	83.52	2.2e+04	349619	! AP002996 Mesorhizobium lo
gb_un:M27461	-	41.00	122.64	148.45	687	! M27461 Figure 3. DNA sequenc
gb_pr:AF172450	-	41.00	116.59	322.63	1676	! AF172450 Homo sapiens opioi
gb_pr:AF172452	-	41.00	114.26	435.06	2363	! AF172452 Homo sapiens opioi

seq_name: gb_ph:AF109874

seq_documentation_block:

```
LOCUS       AF109874    38347 bp    DNA   circular  PHG       15-MAR-2001
DEFINITION  Bacteriophage Tuc2009, complete genome.
ACCESSION   AF109874 L26219 L31348 L31364 L31365 L31366
VERSION     AF109874.2  GI:13346831
KEYWORDS    .
SOURCE      Bacteriophage Tuc2009.
  ORGANISM  Bacteriophage Tuc2009
            Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
REFERENCE   1  (bases 24100 to 24323; 33392 to 34000; 36107 to 38347)
  AUTHORS   Arendt,E.K., Daly,C., Fitzgerald,G.F. and van de Guchte,M.
  TITLE     Molecular characterization of lactococcal bacteriophage Tuc2009 and
            identification and analysis of genes encoding lysin, a putative
            holin, and two structural proteins
  JOURNAL   Appl. Environ. Microbiol. 60 (6), 1875-1883 (1994)
  MEDLINE   94304164
   PUBMED   8031083
REFERENCE   2  (bases 1921 to 2781)
  AUTHORS   van de Guchte,M., Daly,C., Fitzgerald,G.F. and Arendt,E.K.
  TITLE     Identification of the putative repressor-encoding gene cI of the
            temperate lactococcal bacteriophage Tuc2009
  JOURNAL   Gene 144 (1), 93-95 (1994)
  MEDLINE   94299176
   PUBMED   8026765
REFERENCE   3  (bases 37811 to 38347; 1 to 1335)
  AUTHORS   van de Guchte,M., Daly,C., Fitzgerald,G.F. and Arendt,E.K.
  TITLE     Identification of int and attP on the genome of lactococcal
            bacteriophage Tuc2009 and their use for site-specific plasmid
            integration in the chromosome of Tuc2009-resistant Lactococcus
            lactis MG1363
  JOURNAL   Appl. Environ. Microbiol. 60 (7), 2324-2329 (1994)
  MEDLINE   94354646
   PUBMED   8074513
REFERENCE   4  (bases 1 to 38347)
  AUTHORS   van Sinderen,D., van de Guchte,M., Seegers,J.F.M.L. and
            Fitzgerald,G.F.
  TITLE     Molecular analysis of the temperate lactococcal phage Tuc2009
  JOURNAL   Unpublished
REFERENCE   5  (bases 4602 to 11597)
  AUTHORS   McGrath,S., Seegers,J.F.M.L., Fitzgerald,G.F., van Sinderen,D. and
            van de Guchte,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1998) Microbiology, University College Cork,
            College Road, Cork, Ireland
REFERENCE   6  (bases 1 to 38347)
  AUTHORS   van Sinderen,D., van de Guchte,M., Seegers,J.F.M.L. and
            Fitzgerald,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Microbiology, University College Cork,
            College Road, Cork, Ireland
  REMARK    Sequence update by submitter
COMMENT     On Mar 15, 2001 this sequence version replaced gi:508612 gi:496281
            gi:496276 gi:496278 gi:509671 gi:5001694.
FEATURES             Location/Qualifiers
     source          1. .38347
                     /organism="Bacteriophage Tuc2009"
                     /db_xref="taxon:35241"
     CDS             complement(94. .1218)
                     /note="int; orf1"
                     /codon_start=1
                     /transl_table=11
                     /product="integrase"
                     /protein_id="AAA32608.1"
                     /db_xref="GI:508613"
                     /translation="MATYQKRGKTWQYSISRTKQGLPRLTKGGFSTKSDAQAEAMDIE
                     SKLKKGFIVDPIKQEISEYFKDWMELYKKNAIDEMTYKGYEQTLKYLKTYMPNVLISE
```

ITASSYQRALNKFAETHAKASTKGFHTRVRASIQPLIEEGRLQKDFTTRAVVKGNGND
KAEQDKFVNFDEYKQLVDYFRNRLNPNYSSPTMLFIISITSMRASEAFGLVWDDIDFN
NNTIKCRRTWNYRNKVGGFKKPKTDAGIRDIVIDDESMQLLKDFREQQKTLFESLGIK
PIHDFVCYHPYRKIITLSALQNTLDHALKKLNISTPLTVHGLRHTHASVLLYHGVDIM
TVSKRLGHASVAITQQTYIHIIKELENKDKDKIIELLMEL"

CDS complement(1339. .1860)
/note="orf2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK19844.1"
/db_xref="GI:13346832"
/translation="MKNGKTPKAKKPIYKRIWFWIVVVIVVAVIGSALGGGGKGKSGT
STSTSSSSKIKTAEPKKETATPISFEQMNKDYVSNGAAADDKYKGKLLEFQGKVSSVT
ANPFKGTDVTIEAGNFTDNQFQDTKAKVNVNDEMAKQLTSGQTYTFQAKGDGVMMSDG
WVMYLDFNNGVVK"

CDS complement(1921. .2781)
/note="repressor; orf4"
/codon_start=1
/transl_table=11
/product="cI2009"
/protein_id="AAA21825.1"
/db_xref="GI:509672"
/translation="MVIEQINKYVGSKIKDYRKSFGLSQEELAKKIGVGKTTISNYEV
GIRSPKKPQLIKLSEVFDVAIDDFFPQTDSTRMNVSSILSEINKISSQLEEPRQKIVL
NTATNQLDEQNQEKKKESKVIPINKIPDDLPPYISRKILENFIMPSNTMEYEPDEDMV
DVPILGRIAAGLPLDAVENFDGTRPVPAHFLSSARDYYWLMVDGHSMEPKIPYGAYVL
IEAVPDVSDGTIGAVLFQDDCQATLKKVYHEIDCLRLVSINKEFKDQFATQDNPAAVI
GQAVKVEIDL"

CDS 3023. .3253
/note="lambda cro analogue; orf5"
/codon_start=1
/transl_table=11
/product="cro2009"
/protein_id="AAK19845.1"
/db_xref="GI:13346833"
/translation="MAVEKELIALRADEKISRKEMAELIGTTPETYRKKELGESDWWG
AEMFLIASKFNKRIDDIFLDKKSTKSGLEKAS"

CDS 3267. .3983
/note="mp3; orf6"
/codon_start=1
/transl_table=11
/product="major structural protein"
/protein_id="AAK19846.1"
/db_xref="GI:13346834"
/translation="MNQLITITQNENNDQVVSGRELHEFLGVKTRYNDWFEDMVKYGF
TENVDFIGFTEKRVKPQGGRPSVDHALKLDMAKEISMIQRNEKGKQARQYFIEVEKEL
KQQLLPQTPEQQIALLARGNVNLNKKVERIENSVLDLTDRFGLPSNKAKVLQKKVASK
VYMFTGGKYSNAHKKLGAKVFREFYKDLNNRFDVVKYSDIPLSRYDEATEYLDMWQPS
FNTTLEIRGLNSQTSFDFEE"

CDS 3996. .4331
/note="orf7"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK19847.1"
/db_xref="GI:13346835"
/translation="MEQTLEVQATISVLIPEDKILVDKVEYQELKEKDFDGWVGMDVF
TEKSNRSIPTVSKVLRKPDLRKRISVENGGWVYYPNGKGDNWSFRFKEMMDFINKEFY
QKFSGGSGL"

CDS 4328. .4456
/note="orf8"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK19848.1"
/db_xref="GI:13346836"
/translation="MTYTYIVNPETGEILFDLFHDLITQNIRAIKLIAKKLNAVLR"

CDS 4469. .4663
/note="orf9"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK19849.1"
/db_xref="GI:13346837"
/translation="MNKEIEKLANNYKEIINKTSDLALKQNDGDIRKARKWLKEQLFY
TADRATNELIKLSIDNILDY"

CDS 5038. .5253
/note="orf11"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD37096.1"
/db_xref="GI:5001696"
/translation="MYTQIMNGREVLTVPTVIGYKHYDLEKREVVGEVIESTYRRKDG
TMYIIRRSRTEREKAAMLNSCLSDWGY"

CDS 5255. .5353
/note="orf12"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD37097.1"
/db_xref="GI:5001697"
/translation="MSKQQKSPAWHAEQVGNSPKLLLKLYHECLKI"

CDS 5369. .5887
/note="orf13"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD37098.1"
/db_xref="GI:5001698"
/translation="MTEEKEPYKVKNDSELNWVLGKYKEHQIQCDEYEIQAEESRKAI
EEKYNAELYEIEMRRLKLQAEEQKEMDYFKGLAEQYYLTLETKSPKKTINGSVRFSKK
ENVSYDDNLLSELKEKGFGKFITAKTKTTESVDKKALKAFVKDGGQLVSEDGEIVEGF
KFDKTEEFTVKV"

CDS 5896. .6519
/note="orf14"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD37099.1"
/db_xref="GI:5001699"
/translation="MPDKDNSVFELLNSIDVSDKVEKKKSGKNELSYLSWTWAWTEFK
KKFPKATYEIKKFVSKDGNELPYMYDSTTGFMVFTSVTVDDVTHEMWLPVMDGANKAM
KDKPYKYMTKYNGEKSVEQASMFDVNKTIMRCLVKNIAMFGLGLYIYAGEDLPEEPPQ
PQLSDAELIEKYLKQNPENKPNVDEFLKTKSEHEVAEMMKAYIDWSK"

CDS 6519. .6971
/note="SSB; orf15"
/codon_start=1
/transl_table=11
/product="single-strand binding protein"
/protein_id="AAD37100.1"
/db_xref="GI:5001700"
/translation="MINNVVLVGRLTRDPELRHTPQNQAVGTFGLAVNRQFKNANGER
EADLINCVIWRQQAENLAKFAKKGALIGITGRIQTRNYENQQGQKVYVTEVVADTFQM
LESNKTQGQQTSKPQAQNKKPQAPDPFKAPAADPFAGGTEISDDDLPF"

CDS 7096. .7878
/note="Rep; orf16; contains origin of replication"
/codon_start=1
/transl_table=11
/product="replication initiation protein"
/protein_id="AAD37101.1"
/db_xref="GI:5001701"
/translation="MAQRRMFSKEVTTSDLFVDMPSSSQLLYFHLGMEADDEGFIGNA
KMLSRAYGSNNDDLKLLEAKGFVIAFPSGVTVVKDWNLNNKIRKDRQKPTIYTEEKTL

alignment_scores:
Quality: 46.00 Length: 9
Ratio: 5.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-444-281-27 x AF109874/rev ..

Align seg 1/1 to reverse of: AF109874 from: 1 to: 38347

```
      1 HisGluAlaGluProGluAlaGluPro 9
        |||:::|||||||||||||||||||||
  35925 CATAAAGCTGAACCAGAGGCTGAACCT 35899

seq_name: gb_pr:AL391821

seq_documentation_block:
LOCUS       AL391821   158408 bp    DNA             PRI       01-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-212B22 on chromosome X, complete
            sequence.
ACCESSION   AL391821
VERSION     AL391821.7  GI:11121082
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158408)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:10944214.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-212B22 is from the library RPCI-11.1 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-212B22 The true
            left end of clone RP11-487I9 is at 133439 in this sequence. The
            true right end of clone RP11-35G18 is at 153997 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .158408
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-212B22"
                     /clone_lib="RPCI-11.1"
     repeat_region   2. .226
                     /note="L1MB2 repeat: matches 5946. .6163 of consensus"
     repeat_region   295. .525
                     /note="MLT1H repeat: matches 229. .494 of consensus"
     repeat_region   648. .677
                     /note="15 copies 2 mer ca 90% conserved"
     repeat_region   2598. .2649
                     /note="26 copies 2 mer ac 90% conserved"
     repeat_region   2780. .2853
                     /note="37 copies 2 mer ct 74% conserved"
     misc_feature    complement(3446. .3841)
                     /note="match: GSS: Em:AQ136371"
     repeat_region   4491. .4655
                     /note="L1ME repeat: matches 468. .628 of consensus"
     repeat_region   4804. .5008
                     /note="MER58C repeat: matches 3. .88 of consensus"
     repeat_region   5044. .5181
                     /note="L1M4 repeat: matches 1500. .1639 of consensus"
     repeat_region   5271. .5414
                     /note="L1MEc repeat: matches 2242. .2385 of consensus"
     repeat_region   5420. .5718
                     /note="AluSq repeat: matches 1. .298 of consensus"
     repeat_region   5726. .5994
                     /note="AluJo repeat: matches 1. .271 of consensus"
     repeat_region   6011. .6787
                     /note="L1MEc repeat: matches 2356. .2818 of consensus"
     repeat_region   6801. .7262
                     /note="L1PA5 repeat: matches 5682. .6143 of consensus"
     repeat_region   7305. .7750
                     /note="L1MA9 repeat: matches 5860. .6308 of consensus"
     repeat_region   7799. .8106
                     /note="L1M4 repeat: matches 2841. .3151 of consensus"
     repeat_region   8246. .8761
                     /note="L1M4 repeat: matches 869. .1405 of consensus"
     repeat_region   8819. .9030
                     /note="L1M4 repeat: matches 1391. .1596 of consensus"
     repeat_region   9073. .10246
                     /note="L1PA3 repeat: matches 4979. .6146 of consensus"
     repeat_region   10260. .10753
                     /note="L1M4 repeat: matches 1575. .1608 of consensus"
     repeat_region   11391. .11509
                     /note="FLAM_C repeat: matches 1. .132 of consensus"
     repeat_region   11510. .11821
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   11834. .12407
                     /note="L1M4 repeat: matches 2370. .2908 of consensus"
     repeat_region   12408. .12684
                     /note="AluJo repeat: matches 1. .277 of consensus"
     repeat_region   12685. .13112
                     /note="L1M4 repeat: matches 2908. .3346 of consensus"
     repeat_region   13191. .13493
                     /note="L1M4 repeat: matches 3350. .3653 of consensus"
     repeat_region   13515. .13538
                     /note="12 copies 2 mer at 100% conserved"
     repeat_region   13539. .13894
                     /note="L1MA9 repeat: matches 4974. .5334 of consensus"
     repeat_region   13895. .14205
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   14206. .15193
                     /note="L1MA9 repeat: matches 5334. .6305 of consensus"
     repeat_region   15269. .15357
                     /note="L2 repeat: matches 2644. .2748 of consensus"
     repeat_region   15388. .15841
                     /note="L1P repeat: matches 4685. .5138 of consensus"
     misc_feature    complement(15793. .16200)
                     /note="match: GSS: Em:AQ034620"
     repeat_region   16099. .16221
                     /note="L2 repeat: matches 2572. .2710 of consensus"
     repeat_region   16280. .16732
                     /note="L2 repeat: matches 1992. .2403 of consensus"
     repeat_region   17982. .18300
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   18532. .18756
                     /note="MIR repeat: matches 14. .262 of consensus"
     repeat_region   20534. .20587
                     /note="27 copies 2 mer tt 81% conserved"
     repeat_region   21039. .21171
                     /note="AluSg/x repeat: matches 79. .209 of consensus"
     repeat_region   22602. .22902
                     /note="AluSx repeat: matches 2. .302 of consensus"
     repeat_region   23484. .23533
                     /note="25 copies 2 mer gt 96% conserved"
     repeat_region   23532. .23561
                     /note="10 copies 3 mer gtg 90% conserved"
     repeat_region   23877. .25040
                     /note="L2 repeat: matches 1141. .2323 of consensus"
     repeat_region   25257. .25415
```

```
                /note="L2 repeat: matches 2582. .2748 of consensus"
     repeat_region   26553. .26594
                /note="21 copies 2 mer aa 76% conserved"
     misc_feature    26798. .27063
                /note="match: GSS: Em:AQ343061"
     repeat_region   27948. .28106
                /note="MER5B repeat: matches 10. .175 of consensus"
     repeat_region   28430. .28618
                /note="MER53 repeat: matches 1. .189 of consensus"
     repeat_region   28941. .29026
                /note="MIR repeat: matches 112. .198 of consensus"
     repeat_region   30445. .30809
                /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   31228. .31341
                /note="L2 repeat: matches 2629. .2749 of consensus"
     repeat_region   31996. .32079
                /note="L2 repeat: matches 2620. .2698 of consensus"
     repeat_region   32105. .32398
                /note="AluY repeat: matches 1. .294 of consensus"
     misc_feature    complement(32313. .32867)
                /note="match: GSS: Em:AQ357777"
     misc_feature    complement(32386. .32876)
                /note="match: GSS: Em:AQ357649"
     repeat_region   32611. .32654
                /note="22 copies 2 mer tt 79% conserved"
     repeat_region   32804. .32877
                /note="MER5A repeat: matches 21. .114 of consensus"
     repeat_region   33673. .33723
                /note="MIR repeat: matches 102. .152 of consensus"
     misc_feature    complement(34512. .34778)
                /note="match: STS: Em:L24830"
     repeat_region   35370. .35678
                /note="AluY repeat: matches 1. .309 of consensus"
     misc_feature    complement(36574. .36908)
                /note="match: STS: Em:HSC93A6"
     repeat_region   36808. .37545
                /note="L2 repeat: matches 2017. .2749 of consensus"
     repeat_region   37906. .38033
                /note="MER5A repeat: matches 59. .189 of consensus"
     repeat_region   38390. .38732
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   38744. .38995
                /note="MIR repeat: matches 8. .261 of consensus"
     misc_feature    complement(39579. .40134)
                /note="match: STS: Em:HSC18H06"
     repeat_region   39990. .40124
                /note="L2 repeat: matches 2564. .2710 of consensus"
     repeat_region   42729. .42884
                /note="MER5A repeat: matches 4. .184 of consensus"
     repeat_region   43123. .43367
                /note="MIR repeat: matches 3. .261 of consensus"
     repeat_region   44086. .44327
                /note="AluSg/x repeat: matches 70. .308 of consensus"
     repeat_region   45117. .45305
                /note="MER33 repeat: matches 1. .194 of consensus"
     repeat_region   45425. .45951
                /note="L2 repeat: matches 1767. .2358 of consensus"
     repeat_region   46025. .46336
                /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   46474. .46532
                /note="MIR repeat: matches 65. .123 of consensus"

alignment_scores:
           Quality:   44.00          Length:      9
             Ratio:  5.500            Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-444-281-27 x AL391821/rev   ..

 Align seg 1/1  to reverse of: AL391821  from: 1  to: 158408

      1 HisGluAlaGluProGluAlaGluPro 9
```

```
        ||||||||||||||||||    ||||||
   96328 CATGAGGCTGAGCCAGAATTTGAGCCT 96302

seq_name: gb_htg:AC021098

seq_documentation_block:
LOCUS       AC021098   169671 bp    DNA             HTG       01-SEP-2000
DEFINITION  Homo sapiens chromosome X clone RP11-33A2, WORKING DRAFT SEQUENCE,
            18 unordered pieces.
ACCESSION   AC021098
VERSION     AC021098.3  GI:7230836
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169671)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169671)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:6922906.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0033A02
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 81%
            Sequencing vector: plasmid; 19%
            Chemistry: Dye-primer ET; 81% of reads
            Chemistry: Dye-terminator Big Dye; 19% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 159126 bases at least Q40
            Consensus quality: 162659 bases at least Q30
            Consensus quality: 164792 bases at least Q20
            Insert size: 170000; agarose-fp
            Insert size: 167971; sum-of-contigs
            Quality coverage: 4.16 in Q20 bases; agarose-fp
            Quality coverage: 4.21 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1196: contig of 1196 bp in length
            *     1197     1296: gap of unknown length
            *     1297     3284: contig of 1988 bp in length
            *     3285     3384: gap of unknown length
            *     3385     6023: contig of 2639 bp in length
            *     6024     6123: gap of unknown length
            *     6124     8768: contig of 2645 bp in length
            *     8769     8868: gap of unknown length
            *     8869    12171: contig of 3303 bp in length
            *    12172    12271: gap of unknown length
            *    12272    15977: contig of 3706 bp in length
            *    15978    16077: gap of unknown length
            *    16078    24250: contig of 8173 bp in length
            *    24251    24350: gap of unknown length
            *    24351    32681: contig of 8331 bp in length
            *    32682    32781: gap of unknown length
            *    32782    41939: contig of 9158 bp in length
```

```
            *    41940    42039: gap of unknown length
            *    42040    48802: contig of 6763 bp in length
            *    48803    48902: gap of unknown length
            *    48903    57427: contig of 8525 bp in length
            *    57428    57527: gap of unknown length
            *    57528    66323: contig of 8796 bp in length
            *    66324    66423: gap of unknown length
            *    66424    78283: contig of 11860 bp in length
            *    78284    78383: gap of unknown length
            *    78384    93662: contig of 15279 bp in length
            *    93663    93762: gap of unknown length
            *    93763   111526: contig of 17764 bp in length
            *   111527   111626: gap of unknown length
            *   111627   129714: contig of 18088 bp in length
            *   129715   129814: gap of unknown length
            *   129815   144537: contig of 14723 bp in length
            *   144538   144637: gap of unknown length
            *   144638   169671: contig of 25034 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-33A2"
     misc_feature    1. .1196
                     /note="assembly_name:Contig7"
     misc_feature    1297. .3284
                     /note="assembly_name:Contig8"
     misc_feature    3385. .6023
                     /note="assembly_name:Contig9"
     misc_feature    6124. .8768
                     /note="assembly_name:Contig10"
     misc_feature    8869. .12171
                     /note="assembly_name:Contig11"
     misc_feature    12272. .15977
                     /note="assembly_name:Contig12"
     misc_feature    16078. .24250
                     /note="assembly_name:Contig13"
     misc_feature    24351. .32681
                     /note="assembly_name:Contig14"
     misc_feature    32782. .41939
                     /note="assembly_name:Contig15"
     misc_feature    42040. .48802
                     /note="assembly_name:Contig16
                     clone_end:SP6
                     vector_side:right"
     misc_feature    48903. .57427
                     /note="assembly_name:Contig17
                     clone_end:T7
                     vector_side:right"
     misc_feature    57528. .66323
                     /note="assembly_name:Contig18"
     misc_feature    66424. .78283
                     /note="assembly_name:Contig19"
     misc_feature    78384. .93662
                     /note="assembly_name:Contig20"
     misc_feature    93763. .111526
                     /note="assembly_name:Contig21"
     misc_feature    111627. .129714
                     /note="assembly_name:Contig22"
     misc_feature    129815. .144537
                     /note="assembly_name:Contig23"
     misc_feature    144638. .169671
                     /note="assembly_name:Contig24"
BASE COUNT    53919 a   31234 c   31915 g   50898 t    1705 others
ORIGIN


alignment_scores:
            Quality:   44.00              Length:      9
              Ratio:  5.500                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889
```

```
alignment_block:
 US-09-444-281-27 x AC021098   ..

 Align seg 1/1  to: AC021098  from: 1  to: 169671

       1 HisGluAlaGluProGluAlaGluPro 9
         ||||||||||||||||||   ||||||
   44084 CATGAGGCTGAGCCAGAATTTGAGCCT 44110

seq_name: gb_ro:RRPXFPRT

seq_documentation_block:
LOCUS       RRPXFPRT      966 bp    mRNA            ROD       01-OCT-1999
DEFINITION  R.rattus mRNA for PxF protein.
ACCESSION   Y09049
VERSION     Y09049.1  GI:6010290
KEYWORDS    PxF protein.
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   Kammerer,S.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 966)
  AUTHORS   Kammerer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1996) S. Kammerer, Kinderkrankenhaus, Labor f.
            Molekulare Biologie, Lindwurmstr. 4, D- 80337 Muenchen, FRG
FEATURES             Location/Qualifiers
     source          1. .966
                     /organism="Rattus rattus"
                     /db_xref="taxon:10117"
                     /cell_type="myocardiocyte"
     exon            <1. .75
                     /number=1
     gene            join(6. .75,76. .185,186. .351,352. .437,438. .599,600. .776,
                     777. .821,822. .905)
                     /gene="PxF"
     CDS             join(6. .75,76. .185,186. .351,352. .437,438. .599,600. .776,
                     777. .821,822. .905)
                     /gene="PxF"
                     /codon_start=1
                     /product="PxF protein"
                     /protein_id="CAA70258.1"
                     /db_xref="GI:6010291"
                     /translation="MAAAEGGCGAGVEADRELEELLESALDDFDKAKPSPAPSPTISA
                     PDASGPQKRSPGDTAKDALFASQEKFFQELFDSELASQATAEFEKAMKELAEEEPHLV
                     EQFQKLSEAAGRVGSDASSQQEFTSCLKETLSGLAKNATDLQNSGMSEEELTKAMEGL
                     GMDEGDGEGNILPIMQSLMQNLLSKDVLYPSLKEITEKYPEWLQSHQESIPPEQFEKY
                     QQQHSVMGKICEQFEAETPTDSEATHRARFEAVLDLMQQLQDLGHPPKELAGEMPPGL
                     NFDLDALNLSGPPGANGEQCLIM"
     exon            76. .185
                     /gene="PxF"
                     /number=2
     exon            186. .351
                     /gene="PxF"
                     /number=3
     exon            352. .437
                     /gene="PxF"
                     /number=4
     exon            438. .599
                     /gene="PxF"
                     /number=5
     exon            600. .776
                     /gene="PxF"
                     /number=6
     exon            777. .821
                     /gene="PxF"
                     /number=7
     exon            822. .>966
                     /number=8
```

```
BASE COUNT      257 a    247 c    270 g    192 t
ORIGIN

alignment_scores:
            Quality:   43.00             Length:     11
              Ratio:  3.909               Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x RRPXFPRT   ..

 Align seg 1/1  to: RRPXFPRT  from: 1  to: 966

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||:::||||||||||:::|||||||||::::::
     524 CATCATGCAGAGCCTCATGCAGAACCTCTTGTC 556

seq_name: gb_ba:AF127222

seq_documentation_block:
LOCUS       AF127222     2501 bp    DNA             BCT       07-FEB-2000
DEFINITION  Pseudomonas fluorescens outer membrane heme receptor PfhR (pfhR)
            gene, complete cds.
ACCESSION   AF127222
VERSION     AF127222.1  GI:4838476
KEYWORDS    .
SOURCE      Pseudomonas fluorescens.
  ORGANISM  Pseudomonas fluorescens
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 2501)
  AUTHORS   Ochsner,U.A., Johnson,Z. and Vasil,M.L.
  TITLE     Genetics and regulation of two distinct haem-uptake systems, phu
            and has, in Pseudomonas aeruginosa
  JOURNAL   Microbiology 146 (Pt 1), 185-198 (2000)
  MEDLINE   20121752
   PUBMED   10658665
REFERENCE   2  (bases 1 to 2501)
  AUTHORS   Ochsner,U.A. and Vasil,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1999) Microbiology, University of Colorado Health
            Sci Ctr, 4200 E. Ninth Avenue, Denver, CO 80262, USA
FEATURES             Location/Qualifiers
     source          1. .2501
                     /organism="Pseudomonas fluorescens"
                     /strain="ATCC 15453"
                     /db_xref="taxon:294"
                     /db_xref="ATCC:15453"
     gene            195. .2465
                     /gene="pfhR"
     CDS             195. .2465
                     /gene="pfhR"
                     /function="heme uptake"
                     /codon_start=1
                     /transl_table=11
                     /product="outer membrane heme receptor PfhR"
                     /protein_id="AAD31012.1"
                     /db_xref="GI:4838477"
                     /translation="MGTNLSYRKLGTNNIVLEKRSAGSALALQQVTISATRNEQDVNS
                     VPSTVSVHDRAELDRQHVNNIRELVRYEPGVSVGGAGTRSGNAGYNIRGIDGDRVLTQ
                     VDGVEVPDNFFNGPYAKTRRNYVDPEIVKRVEILRGPASALYGSSAIGGAVSYFTLDP
                     DDIIKPGQDVGARLKTGYSSADESWLTSGTFAGRVQDFDGLLHLSQRNGHETESYDGN
                     NATGLARTGANPEDARTNVLAKLGWNYGDDNRLGLTYEKFKDDRDVNLKNAVGGPFIG
                     GRGMNLYRDRRGNDTITRERFGLENTFALESPIADRIKTSLNYQIAKTDQTTAEIYQA
                     GRRVLRTRDTLYEEKQWVFDAQLDKAFSLGETDHQVTYGTTLKQQKVTGSREGSASCL
                     AIGAGCTAIGAPSPSASDSVKKASDFPDPTINTYSLFAQDQITWDKWTFLPAVRYDYT
                     RLKPKLTQEFLNTVNPTGAYTVDGKDKTWNRVTPKFGLTYALTDNYTWFGQYAEGFRT
                     PSAKALYGRFENLNLGYTVEPNPDLKPETSKGIETGIRGKFDEGSFDIAVYYNKYRDF
                     IDEDKPVAGGTVEQFQAVNIKRATIKGVEAKGRLNLDTLGAPKGLYTQGSVAYTYGRN
                     DDNGEPLNSVNPLKGVFGLGYDQDNYGGLVSWTVVKKQNRVDSTTFHAPDGGTDGPFK
                     TPGFGILDLTAYYKVSKDVTVNGGLYNLTDKKYWNWDDVRSYDSVGEAGVTGPANLDR
                     LTQPGRNFAINVIWDI"
```

```
BASE COUNT      561 a    874 c    698 g    368 t
ORIGIN

alignment_scores:
            Quality:   43.00             Length:     10
              Ratio:  4.778               Gaps:      0
 Percent Similarity:  90.000  Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x AF127222/rev   ..

 Align seg 1/1  to reverse of: AF127222  from: 1  to: 2501

       1 HisGluAlaGluProGluAlaGluProIle 10
         ||||||||||||||||:::   ||||||:::
    1366 CACGAGGCCGAGCCTTCGCGTGAGCCGGTG 1337

seq_name: gb_htg:AC015790

seq_documentation_block:
LOCUS       AC015790    79795 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-2L12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC015790
VERSION     AC015790.2  GI:9109862
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-2L12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6446897.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1552
                Center clone name: 2_L_12
            --------------.
            * NOTE: This record contains 80 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
```

```
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*       1      901: contig of 901 bp in length
*     902 1001: gap of      100 bp
*    1002     1910: contig of 909 bp in length
*    1911 2010: gap of      100 bp
*    2011     2938: contig of 928 bp in length
*    2939 3038: gap of      100 bp
*    3039     3933: contig of 895 bp in length
*    3934 4033: gap of      100 bp
*    4034     4931: contig of 898 bp in length
*    4932 5031: gap of      100 bp
*    5032     5905: contig of 874 bp in length
*    5906 6005: gap of      100 bp
*    6006     6894: contig of 889 bp in length
*    6895 6994: gap of      100 bp
*    6995     7902: contig of 908 bp in length
*    7903 8002: gap of      100 bp
*    8003     8887: contig of 885 bp in length
*    8888 8987: gap of      100 bp
*    8988     9879: contig of 892 bp in length
*    9880 9979: gap of      100 bp
*    9980    10888: contig of 909 bp in length
*   10889 10988: gap of      100 bp
*   10989    11908: contig of 920 bp in length
*   11909 12008: gap of      100 bp
*   12009    12888: contig of 880 bp in length
*   12889 12988: gap of      100 bp
*   12989    13882: contig of 894 bp in length
*   13883 13982: gap of      100 bp
*   13983    14889: contig of 907 bp in length
*   14890 14989: gap of      100 bp
*   14990    15897: contig of 908 bp in length
*   15898 15997: gap of      100 bp
*   15998    16890: contig of 893 bp in length
*   16891 16990: gap of      100 bp
*   16991    17891: contig of 901 bp in length
*   17892 17991: gap of      100 bp
*   17992    18891: contig of 900 bp in length
*   18892 18991: gap of      100 bp
*   18992    19894: contig of 903 bp in length
*   19895 19994: gap of      100 bp
*   19995    20860: contig of 866 bp in length
*   20861 20960: gap of      100 bp
*   20961    21867: contig of 907 bp in length
*   21868 21967: gap of      100 bp
*   21968    22868: contig of 901 bp in length
*   22869 22968: gap of      100 bp
*   22969    23859: contig of 891 bp in length
*   23860 23959: gap of      100 bp
*   23960    24832: contig of 873 bp in length
*   24833 24932: gap of      100 bp
*   24933    25833: contig of 901 bp in length
*   25834 25933: gap of      100 bp
*   25934    26845: contig of 912 bp in length
*   26846 26945: gap of      100 bp
*   26946    27854: contig of 909 bp in length
*   27855 27954: gap of      100 bp
*   27955    28864: contig of 910 bp in length
*   28865 28964: gap of      100 bp
*   28965    29856: contig of 892 bp in length
*   29857 29956: gap of      100 bp
*   29957    30835: contig of 879 bp in length
*   30836 30935: gap of      100 bp
*   30936    31843: contig of 908 bp in length
*   31844 31943: gap of      100 bp
*   31944    32864: contig of 921 bp in length
*   32865 32964: gap of      100 bp
*   32965    33849: contig of 885 bp in length
*   33850 33949: gap of      100 bp
*   33950    34823: contig of 874 bp in length
*   34824 34923: gap of      100 bp
*   34924    35815: contig of 892 bp in length
*   35816 35915: gap of      100 bp
*   35916    36826: contig of 911 bp in length
*   36827 36926: gap of      100 bp
*   36927    37836: contig of 910 bp in length
*   37837 37936: gap of      100 bp
*   37937    38837: contig of 901 bp in length
*   38838 38937: gap of      100 bp
*   38938    39818: contig of 881 bp in length
*   39819 39918: gap of      100 bp
*   39919    40829: contig of 911 bp in length
*   40830 40929: gap of      100 bp
*   40930    41819: contig of 890 bp in length
*   41820 41919: gap of      100 bp
*   41920    42789: contig of 870 bp in length
*   42790 42889: gap of      100 bp
*   42890    43767: contig of 878 bp in length
*   43768 43867: gap of      100 bp
*   43868    44753: contig of 886 bp in length
*   44754 44853: gap of      100 bp
*   44854    45777: contig of 924 bp in length
*   45778 45877: gap of      100 bp
*   45878    46780: contig of 903 bp in length
*   46781 46880: gap of      100 bp
*   46881    47802: contig of 922 bp in length
*   47803 47902: gap of      100 bp
*   47903    48803: contig of 901 bp in length
*   48804 48903: gap of      100 bp
*   48904    49807: contig of 904 bp in length
*   49808 49907: gap of      100 bp
*   49908    50790: contig of 883 bp in length
*   50791 50890: gap of      100 bp
*   50891    51810: contig of 920 bp in length
*   51811 51910: gap of      100 bp
*   51911    52812: contig of 902 bp in length
*   52813 52912: gap of      100 bp
*   52913    53796: contig of 884 bp in length
*   53797 53896: gap of      100 bp
*   53897    54818: contig of 922 bp in length
*   54819 54918: gap of      100 bp
*   54919    55819: contig of 901 bp in length
*   55820 55919: gap of      100 bp
*   55920    56836: contig of 917 bp in length
*   56837 56936: gap of      100 bp
*   56937    57850: contig of 914 bp in length
*   57851 57950: gap of      100 bp
*   57951    58874: contig of 924 bp in length
*   58875 58974: gap of      100 bp
*   58975    59847: contig of 873 bp in length
*   59848 59947: gap of      100 bp
*   59948    60838: contig of 891 bp in length
*   60839 60938: gap of      100 bp
*   60939    61829: contig of 891 bp in length
*   61830 61929: gap of      100 bp
*   61930    62806: contig of 877 bp in length
*   62807 62906: gap of      100 bp
*   62907    63784: contig of 878 bp in length
*   63785 63884: gap of      100 bp
*   63885    64792: contig of 908 bp in length
*   64793 64892: gap of      100 bp
*   64893    65789: contig of 897 bp in length
*   65790 65889: gap of      100 bp
*   65890    66771: contig of 882 bp in length
*   66772 66871: gap of      100 bp
*   66872    67778: contig of 907 bp in length
*   67779 67878: gap of      100 bp
*   67879    68780: contig of 902 bp in length
*   68781 68880: gap of      100 bp
*   68881    69764: contig of 884 bp in length
*   69765 69864: gap of      100 bp
*   69865    70743: contig of 879 bp in length
*   70744 70843: gap of      100 bp
*   70844    71756: contig of 913 bp in length
```

```
          *    71757 71856: gap of      100 bp
          *    71857    72761: contig of 905 bp in length

alignment_scores:
           Quality:   43.00            Length:     10
             Ratio:  4.300              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-27 x AC015790/rev   ..

 Align seg 1/1  to reverse of: AC015790  from: 1  to: 79795

       2 GluAlaGluProGluAlaGluProIleMet 11
         |||:::::|||||||||||||||||||||||
    4914 GAAAGCTCCCCAGAAGCAGAACCCATCATG 4885

seq_name: gb_pr:AL138709

seq_documentation_block:
LOCUS       AL138709    96874 bp    DNA             PRI       08-AUG-2001
DEFINITION  Human DNA sequence from clone RP11-460G11 on chromosome 13,
            complete sequence.
ACCESSION   AL138709
VERSION     AL138709.19  GI:15141998
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 96874)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 9, 2001 this sequence version replaced gi:14626037.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-460G11 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-460G11 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-460G11 is at 96874 in this
            sequence. The true right end of clone RP11-335G18 is at 2000 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .96874
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-460G11"
                     /clone_lib="RPCI-11.2"
     repeat_region   638. .775
                     /note="AluSp/q repeat: matches 159. .296 of consensus"
     repeat_region   2734. .2953
                     /note="MIR repeat: matches 20. .252 of consensus"
     repeat_region   3412. .3570
                     /note="Charlie4 repeat: matches 1740. .1907 of consensus"
     repeat_region   4223. .4292
                     /note="MIR repeat: matches 133. .205 of consensus"
     repeat_region   4354. .4475
                     /note="MER5B repeat: matches 1. .129 of consensus"
     repeat_region   4980. .5129
                     /note="MER5A repeat: matches 13. .164 of consensus"
     repeat_region   5146. .5284
                     /note="MIR repeat: matches 9. .148 of consensus"
     repeat_region   5314. .5483
                     /note="Charlie4a repeat: matches 17. .174 of consensus"
     repeat_region   5484. .5788
                     /note="AluY repeat: matches 5. .309 of consensus"
     repeat_region   5789. .6029
                     /note="Charlie4a repeat: matches 174. .457 of consensus"
     repeat_region   6536. .6602
                     /note="L2 repeat: matches 2675. .2743 of consensus"
     repeat_region   6719. .7125
                     /note="MLT2FB repeat: matches 4. .414 of consensus"
     repeat_region   7364. .7568
                     /note="MIR repeat: matches 37. .251 of consensus"
     repeat_region   10015. .10064
                     /note="MER3 repeat: matches 41. .90 of consensus"
     repeat_region   10027. .10085
                     /note="MER3 repeat: matches 144. .204 of consensus"
     repeat_region   10086. .10325
                     /note="L1MA1 repeat: matches 5550. .5792 of consensus"
     repeat_region   10326. .10634
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   10635. .11160
                     /note="L1MA1 repeat: matches 5792. .6304 of consensus"
     repeat_region   11161. .11179
                     /note="MER3 repeat: matches 130. .144 of consensus"
     repeat_region   11182. .11347
                     /note="MER3 repeat: matches 1. .167 of consensus"
     repeat_region   11713. .11780
                     /note="4 copies 17 mer 86% conserved"
     repeat_region   11787. .11881
                     /note="L2 repeat: matches 2575. .2699 of consensus"
     repeat_region   12183. .12292
                     /note="MIR repeat: matches 126. .246 of consensus"
     repeat_region   13008. .13053
                     /note="23 copies 2 mer tc 73% conserved"
     repeat_region   13063. .13218
                     /note="6 copies 26 mer 67% conserved"
     repeat_region   13065. .13178
                     /note="57 copies 2 mer tc 67% conserved"
     repeat_region   13674. .13759
                     /note="MIR repeat: matches 78. .165 of consensus"
     repeat_region   15493. .15544
                     /note="26 copies 2 mer ac 90% conserved"
     repeat_region   16508. .17150
                     /note="MER41B repeat: matches 3. .635 of consensus"
     repeat_region   18228. .18286
                     /note="tRNA-Ile-ATT repeat: matches 14. .74 of consensus"
     repeat_region   18765. .19215
                     /note="L2 repeat: matches 2290. .2748 of consensus"
     repeat_region   19216. .19985
                     /note="LTR39 repeat: matches 1. .794 of consensus"
     repeat_region   19986. .20070
                     /note="L2 repeat: matches 2208. .2290 of consensus"
     repeat_region   21039. .21339
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   21350. .21658
                     /note="AluSq repeat: matches 1. .307 of consensus"
```

```
repeat_region   21674. .21731
                /note="MER53 repeat: matches 127. .186 of consensus"
repeat_region   21732. .22043
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   22044. .22175
                /note="MER53 repeat: matches 6. .127 of consensus"
repeat_region   22862. .23024
                /note="MIR repeat: matches 87. .248 of consensus"
repeat_region   24844. .25143
                /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region   26057. .26200
                /note="MER5A repeat: matches 17. .187 of consensus"
repeat_region   27843. .28278
                /note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region   28375. .28436
                /note="L2 repeat: matches 2645. .2710 of consensus"
repeat_region   30397. .30480
                /note="2 copies 42 mer 90% conserved"
repeat_region   31856. .32054
                /note="MIR repeat: matches 37. .261 of consensus"
repeat_region   32358. .32528
                /note="L2 repeat: matches 2092. .2279 of consensus"
repeat_region   32608. .32648
                /note="MLT1D repeat: matches 201. .240 of consensus"
repeat_region   32649. .32947
                /note="AluJo repeat: matches 1. .294 of consensus"
repeat_region   32948. .33144
                /note="MLT1D repeat: matches 1. .201 of consensus"
repeat_region   33183. .33230
                /note="L2 repeat: matches 2677. .2724 of consensus"
repeat_region   34446. .34647
                /note="MER2 repeat: matches 1. .211 of consensus"
repeat_region   34678. .34757
                /note="MER2 repeat: matches 264. .344 of consensus"
repeat_region   38187. .38646
                /note="MLT1C repeat: matches 6. .466 of consensus"
repeat_region   38848. .38958
                /note="MER81 repeat: matches 3. .114 of consensus"
repeat_region   40042. .40152
                /note="L2 repeat: matches 2381. .2492 of consensus"
repeat_region   41116. .41251
                /note="MIR repeat: matches 15. .151 of consensus"
repeat_region   42941. .43367
                /note="LTR16B repeat: matches 39. .464 of consensus"
repeat_region   44219. .44302
                /note="MLT1J repeat: matches 242. .318 of consensus"
repeat_region   46089. .47392
                /note="L1PA2 repeat: matches 3642. .4945 of consensus"
repeat_region   47393. .48587
                /note="L1PA2 repeat: matches 4952. .6146 of consensus"
repeat_region   48647. .48781
                /note="MIR repeat: matches 85. .234 of consensus"
repeat_region   49096. .49396
                /note="AluSg repeat: matches 1. .296 of consensus"
repeat_region   49623. .51400
                /note="L1PA7 repeat: matches 4358. .6143 of consensus"
repeat_region   52702. .53151
                /note="MER39 repeat: matches 13. .455 of consensus"
repeat_region   53233. .53451
                /note="MER72 repeat: matches 445. .668 of consensus"
repeat_region   53452. .53750
                /note="AluSq repeat: matches 1. .294 of consensus"
repeat_region   53751. .54215
                /note="MER72 repeat: matches 1. .445 of consensus"
repeat_region   54223. .54364
                /note="L2 repeat: matches 2563. .2750 of consensus"
repeat_region   54866. .55171
                /note="AluY repeat: matches 2. .303 of consensus"
repeat_region   55239. .55292
                /note="27 copies 2 mer tt 70% conserved"
repeat_region   57830. .58112
                /note="AluSq repeat: matches 3. .304 of consensus"
repeat_region   58224. .58516
                /note="MLT1I repeat: matches 98. .397 of consensus"

alignment_scores:
            Quality:   43.00            Length:     11
              Ratio:  4.778              Gaps:      0
 Percent Similarity:  81.818   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x AL138709/rev   ..

 Align seg 1/1  to reverse of: AL138709  from: 1  to: 96874

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         ||||||   |||||||||   ||||||::::::
   66249 CACGAGCCAGAGCCTGAGCCCGAGCCCCTGGTT 66217

seq_name: gb_htg:AC019290

seq_documentation_block:
LOCUS       AC019290   176712 bp    DNA             HTG       09-SEP-2000
DEFINITION  Homo sapiens clone RP11-649M10, WORKING DRAFT SEQUENCE, 15
            unordered pieces.
ACCESSION   AC019290
VERSION     AC019290.4  GI:10045469
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176712)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-649M10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176712)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA .
COMMENT     On Sep 9, 2000 this sequence version replaced gi:6721321.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4312
                Center clone name: 649_M_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167161 bases at least Q40
                Consensus quality: 171786 bases at least Q30
                Consensus quality: 173554 bases at least Q20
```

```
            Insert size: 182000; agarose-fp
            Insert size: 175312; sum-of-contigs
            Quality coverage: 4.3 in Q20 bases; agarose-fp
            Quality coverage: 4.5 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    31590: contig of 31590 bp in length
            *    31591 31690: gap of      100 bp
            *    31691    33087: contig of 1397 bp in length
            *    33088 33187: gap of      100 bp
            *    33188    34887: contig of 1700 bp in length
            *    34888 34987: gap of      100 bp
            *    34988    37644: contig of 2657 bp in length
            *    37645 37744: gap of      100 bp
            *    37745    40979: contig of 3235 bp in length
            *    40980 41079: gap of      100 bp
            *    41080    46341: contig of 5262 bp in length
            *    46342 46441: gap of      100 bp
            *    46442    53992: contig of 7551 bp in length
            *    53993 54092: gap of      100 bp
            *    54093    60116: contig of 6024 bp in length
            *    60117 60216: gap of      100 bp
            *    60217    68183: contig of 7967 bp in length
            *    68184 68283: gap of      100 bp
            *    68284    76928: contig of 8645 bp in length
            *    76929 77028: gap of      100 bp
            *    77029    95627: contig of 18599 bp in length
            *    95628 95727: gap of      100 bp
            *    95728   115794: contig of 20067 bp in length
            *   115795 115894: gap of      100 bp
            *   115895   143115: contig of 27221 bp in length
            *   143116 143215: gap of      100 bp
            *   143216   174379: contig of 31164 bp in length
            *   174380 174479: gap of      100 bp
            *   174480   176712: contig of 2233 bp in length.
FEATURES             Location/Qualifiers.
     source          1. .176712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-649M10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .31590
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    31691. .33087
                     /note="assembly_fragment"
     misc_feature    33188. .34887
                     /note="assembly_fragment"
     misc_feature    34988. .37644
                     /note="assembly_fragment"
     misc_feature    37745. .40979
                     /note="assembly_fragment"
     misc_feature    41080. .46341
                     /note="assembly_fragment"
     misc_feature    46442. .53992
                     /note="assembly_fragment"
     misc_feature    54093. .60116
                     /note="assembly_fragment"
     misc_feature    60217. .68183
                     /note="assembly_fragment"
     misc_feature    68284. .76928
                     /note="assembly_fragment"
     misc_feature    77029. .95627
                     /note="assembly_fragment"
     misc_feature    95728. .115794
                     /note="assembly_fragment"
     misc_feature    115895. .143115
                     /note="assembly_fragment"
     misc_feature    143216. .174379
                     /note="assembly_fragment"
     misc_feature    174480. .176712
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    53093 a  35421 c  35789 g  51008 t   1401 others
ORIGIN


alignment_scores:
            Quality:   43.00            Length:     11
              Ratio:  4.300              Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 72.727

alignment_block:
 US-09-444-281-27 x AC019290/rev   ..

 Align seg 1/1  to reverse of: AC019290  from: 1  to: 176712

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         |||||||||:::||||||||||:::   ||||||
   65329 CATGAGGCCTCCCCAGAAGCAGATGCCATCATG 65297

seq_name: gb_htg:AC016270

seq_documentation_block:
LOCUS        AC016270   189022 bp    DNA             HTG       16-MAR-2000
DEFINITION   Homo sapiens clone RP11-17I11, WORKING DRAFT SEQUENCE, 15 unordered
             pieces.
ACCESSION    AC016270
VERSION      AC016270.4  GI:7248975
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 189022)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens, clone RP11-17I11
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 189022)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
             Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
             Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
             Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
             McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
             Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
             Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
             Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Mar 16, 2000 this sequence version replaced gi:6970443.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ---------- Genome Center
                 Center: Whitehead Institute/ MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www-seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
             ---------- Project Information
                 Center project name: L3645
```

```
           Center clone name: 17_I_11
          ------------- Summary Statistics
           Sequencing vector: M13; M77815; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 182484 bases at least Q40
           Consensus quality: 184881 bases at least Q30
           Consensus quality: 186037 bases at least Q20
           Insert size: 186000; agarose-fp
           Insert size: 187622; sum-of-contigs
           Quality coverage: 5.8 in Q20 bases; agarose-fp
           Quality coverage: 5.8 in Q20 bases; sum-of-contigs
          --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 15 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1585: contig of 1585 bp in length
          *     1586 1685: gap of      100 bp
          *     1686     3061: contig of 1376 bp in length
          *     3062 3161: gap of      100 bp
          *     3162     7504: contig of 4343 bp in length
          *     7505 7604: gap of      100 bp
          *     7605    10536: contig of 2932 bp in length
          *    10537 10636: gap of      100 bp
          *    10637    13824: contig of 3188 bp in length
          *    13825 13924: gap of      100 bp
          *    13925    18777: contig of 4853 bp in length
          *    18778 18877: gap of      100 bp
          *    18878    29087: contig of 10210 bp in length
          *    29088 29187: gap of      100 bp
          *    29188    40374: contig of 11187 bp in length
          *    40375 40474: gap of      100 bp
          *    40475    55299: contig of 14825 bp in length
          *    55300 55399: gap of      100 bp
          *    55400    67577: contig of 12178 bp in length
          *    67578 67677: gap of      100 bp
          *    67678    86116: contig of 18439 bp in length
          *    86117 86216: gap of      100 bp
          *    86217   104553: contig of 18337 bp in length
          *   104554 104653: gap of      100 bp
          *   104654   125313: contig of 20660 bp in length
          *   125314 125413: gap of      100 bp
          *   125414   154037: contig of 28624 bp in length
          *   154038 154137: gap of      100 bp
          *   154138   189022: contig of 34885 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189022
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-17I11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1585
                     /note="assembly_fragment"
     misc_feature    1686. .3061
                     /note="assembly_fragment"
     misc_feature    3162. .7504
                     /note="assembly_fragment"
     misc_feature    7605. .10536
                     /note="assembly_fragment"
     misc_feature    10637. .13824
                     /note="assembly_fragment"
     misc_feature    13925. .18777
                     /note="assembly_fragment"
     misc_feature    18878. .29087
                     /note="assembly_fragment"
     misc_feature    29188. .40374
                     /note="assembly_fragment
                     clone_end:SP6"
     misc_feature    40475. .55299
                     /note="assembly_fragment"
     misc_feature    55400. .67577
                     /note="assembly_fragment"
     misc_feature    67678. .86116
                     /note="assembly_fragment"
     misc_feature    86217. .104553
                     /note="assembly_fragment"
     misc_feature    104654. .125313
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    125414. .154037
                     /note="assembly_fragment"
     misc_feature    154138. .189022
                     /note="assembly_fragment"
BASE COUNT    55798 a  38604 c  38373 g  54845 t   1402 others
ORIGIN


alignment_scores:
            Quality:   43.00              Length:     11
              Ratio:  4.778                Gaps:      0
 Percent Similarity:  81.818   Percent Identity: 63.636

alignment_block:
 US-09-444-281-27 x AC016270   ..

 Align seg 1/1  to: AC016270  from: 1  to: 189022

       1 HisGluAlaGluProGluAlaGluProIleMet 11
         ||||||   |||||||||   ||||||::::::
   43795 CACGAGCCAGAGCCTGAGCCCGAGCCCCTGGTT 43827

seq_name: gb_in:AMAPID22

seq_documentation_block:
LOCUS       AMAPID22      607 bp    mRNA            INV       03-MAY-1993
DEFINITION  A.mellifera Apid22 mRNA.
ACCESSION   X72576
VERSION     X72576.1  GI:297064
KEYWORDS    antibiotic peptide; apid22 gene; Apidaecin; precursor protein.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
            Aculeata; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Casteels-Josson,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1993) K. Casteels-Josson, Memorial Sloan
            Kettering Cancer Center, 1275 York Avenue, New York 10021, NY, USA
REFERENCE   2  (bases 1 to 607)
  AUTHORS   Casteels-Josson,K., Capaci,T., Casteels,P. and Tempst,P.
  TITLE     Apidaecin multipeptide precursor structure: a putative mechanism
            for amplification of the insect antibacterial response
  JOURNAL   EMBO J. 12 (4), 1569-1578 (1993)
  MEDLINE   93223697
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Apis mellifera"
                     /db_xref="taxon:7460"
     primer_bind     1. .14
     sig_peptide     20. .76
                     /gene="Apid22"
                     /product="secretion signal"
     CDS             20. .454
                     /gene="Apid22"
                     /codon_start=1
                     /product="Apidaecin precursor"
                     /protein_id="CAA51168.1"
                     /db_xref="GI:297065"
```

```
                     /db_xref="SWISS-PROT:P35581"
                     /translation="MKNFALAILVVTFVVAVFGNTNLDPPTRPTRLRREAEPEAEPGN
                     NRPVYIPQPRPPHPRLRREAEPEAEPGNNRPVYIPQPRPPHPRLRREAEPEAEPGNNR
                     PVYIPQPRPPHPRLRREAEPEAEPGNNRPVYIPQPRPPHPRI"
     gene            20. .454
                     /gene="Apid22"
     mat_peptide     77. .451
                     /gene="Apid22"
                     /note="promultipeptide"
                     /product="Apidaecin"
     misc_signal     454. .458
                     /note="ATTTA"
     misc_signal     490. .494
                     /note="ATTTA"
     polyA_signal    494. .499
     polyA_signal    550. .555
     primer_bind     602. .607
BASE COUNT      202 a    149 c    106 g    150 t
ORIGIN


alignment_scores:
            Quality:   42.00             Length:      8
              Ratio:  5.250               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-27 x AMAPID22   ..

 Align seg 1/1  to: AMAPID22  from: 1  to: 607

       2 GluAlaGluProGluAlaGluPro 9
         |||||||||||||||||||||||||
     122 GAAGCTGAACCGGAAGCTGAACCC 145

seq_name: gb_pl:AF034945

seq_documentation_block:
LOCUS       AF034945      790 bp    mRNA            PLN       10-DEC-1997
DEFINITION  Zea mays glycine-rich RNA binding protein (GRP) mRNA, complete cds.
ACCESSION   AF034945
VERSION     AF034945.1  GI:2668741
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   Baysdorfer,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Biological Sciences, California State
            University, Hayward, CA 94542, USA
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Zea mays"
                     /strain="B73"
                     /db_xref="taxon:4577"
     gene            1. .790
                     /gene="GRP"
     CDS             72. .509
                     /gene="GRP"
                     /codon_start=1
                     /product="glycine-rich RNA binding protein"
                     /protein_id="AAB88616.1"
                     /db_xref="GI:2668742"
                     /translation="MAASDVEYRCFVGGLAWATDDHSLNNAFSTYGEVLESKIILDRE
                     TQRSRGFGFVTFSTEDAMRSAIEGMNGKELDGRNITVNEAQSRGGRGGGGGGYGGGRR
                     DGGGYGGGGGGYGGGRGGYGGGGGYGGANRGGGYGNNDGNWRN"
BASE COUNT      141 a    211 c    262 g    176 t
ORIGIN
```

```
alignment_scores:
            Quality:   42.00             Length:     10
              Ratio:  5.250               Gaps:      0
 Percent Similarity:  80.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x AF034945/rev   ..

 Align seg 1/1  to reverse of: AF034945  from: 1  to: 790

       1 HisGluAlaGluProGluAlaGluProIle 10
         ||||||   |||||||||   ||||||:::
      56 CACGAACCAGAACCTGAACACGAACCGGTG 27

seq_name: gb_ro:MMY12380

seq_documentation_block:
LOCUS       MMY12380      915 bp    DNA             ROD       31-JUL-1998
DEFINITION  M.musculus ITGA7 gene, promoter and exon 1 (and joined CDS).
ACCESSION   Y12380
VERSION     Y12380.1  GI:3378242
KEYWORDS    integrin alpha 7; ITGA7 gene; laminin receptor.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   Saher,G., Echtermeyer,F., Beier,D.R., Poeschl,E. and Mayer,U.
  TITLE     Genomic organization and chromosomal localization of the mouse
            integrin alpha7 gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 915)
  AUTHORS   Saher,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-1997) G. Saher, Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152, Martinsried, FRG
COMMENT     Related sequences L23421, L23422 & L23423.
FEATURES             Location/Qualifiers
     source          1. .915
                     /organism="Mus musculus"
                     /strain="129Sv"
                     /db_xref="taxon:10090"
                     /chromosome="10"
     promoter        1. .389
     mRNA            join(390. .778,Y12383.1:79. .206,Y12383.1:307. .386,
                     Y12384.1:87. .342,Y12385.1:14. .145,Y12385.1:391. .510,
                     Y12386.1:113. .320,Y12386.1:423. .616,Y12386.1:859. .947,
                     Y12386.1:1057. .1184,Y12386.1:1311. .1406,
                     Y12386.1:1547. .1608,Y12386.1:1710. .1879,Y12387.1:110. .259,
                     Y12388.1:110. .225,Y12388.1:622. .814,Y12389.1:109. .189,
                     Y12389.1:276. .355,Y12389.1:454. .528,Y12389.1:667. .769,
                     Y12389.1:1256. .1432,Y12389.1:1597. .1725,Y12390.1:64. .177,
                     Y12390.1:381. .479,Y12390.1:567. .692,Y12381.1:106. .218,
                     Y12382.1:53. .804)
                     /note="A variant"
                     /label=ITGRNAA
     mRNA            join(390. .778,Y12383.1:79. .206,Y12383.1:307. .386,
                     Y12384.1:87. .342,Y12385.1:14. .145,Y12385.1:391. .510,
                     Y12386.1:113. .320,Y12386.1:423. .616,Y12386.1:859. .947,
                     Y12386.1:1057. .1184,Y12386.1:1311. .1406,
                     Y12386.1:1547. .1608,Y12386.1:1710. .1879,Y12387.1:110. .259,
                     Y12388.1:110. .225,Y12388.1:622. .814,Y12389.1:109. .189,
                     Y12389.1:276. .355,Y12389.1:454. .528,Y12389.1:667. .769,
                     Y12389.1:1256. .1432,Y12389.1:1597. .1725,Y12390.1:64. .177,
                     Y12390.1:381. .479,Y12390.1:567. .692,Y12382.1:53. .804)
                     /note="B variant"
                     /label=ITGRNAB
     exon            390. .778
                     /label=ex1
                     /number=1
     misc_feature    573. .778
                     /gene="ITGA7"
                     /note="start codon"
```

```
                /label=start
gene            join(57[illegible]. .778,Y12383.1:79. .386,Y12384.1:87. .342,
                Y12385.1:14. .510,Y12386.1:113. .1879,Y12387.1:110. .259,
                Y12388.1:110. .814,Y12389.1:109. .1725,Y12390.1:64. .692,
                Y12381.1:106. .218,Y12382.1:53. .283)
                /gene="ITGA7"
CDS             join(573. .778,Y12383.1:79. .206,Y12383.1:307. .386,
                Y12384.1:87. .342,Y12385.1:14. .145,Y12385.1:391. .510,
                Y12386.1:113. .320,Y12386.1:423. .616,Y12386.1:859. .947,
                Y12386.1:1057. .1184,Y12386.1:1311. .1406,
                Y12386.1:1547. .1608,Y12386.1:1710. .1879,Y12387.1:110. .259,
                Y12388.1:110. .225,Y12388.1:622. .814,Y12389.1:109. .189,
                Y12389.1:276. .355,Y12389.1:454. .528,Y12389.1:667. .769,
                Y12389.1:1256. .1432,Y12389.1:1597. .1725,Y12390.1:64. .177,
                Y12390.1:381. .479,Y12390.1:567. .692,Y12382.1:53. .283)
                /gene="ITGA7"
                /function="laminin receptor"
                /note="B variant"
                /codon_start=1
                /label=ITGCDSB
                /product="integrin alpha 7"
                /protein_id="CAA73023.1"
                /db_xref="GI:3378244"
                /db_xref="SPTREMBL:O88732"
                /translation="MARIPRCDFLRPPGIYYLITSLLAGLFLPPAIAFNLDVMGAIRK
                EGEPGSLFGFSVALHRQLQPRPQSWLLVGAPQALALPGQQANRTGGLFACPLSLEETD
                CYRVDIDRGANVQKESKENQWLGVSVRSQGPGGKIVTCAHRYESRQRVDQALETRDVI
                GRCFVLSQDLAIRDELDGGEWKFCEGRPQGHEQFGFCQQGTAATFSPDSHYLVFGAPG
                TYNWKGTARVELCAQGSPDLAHLDDGPYEAGGEKEQDPRLIPVPANSYLGLLFVTNID
                SSDPDQLVYKTLDPADRLTGPAGDLTLNSYLGFSIDSGKGLMRSEELSFVAGAPRANH
                KGAVVILRKDSASRLIPEVVLSGERLTSGFGYSLAVTDLNNDGWADLIVGAPYFFERQ
                EELGGAVYVYMNQGGHWADISPLRICGSPDSMFGISLAVLGDLNQDGFPDIAVGAPFD
                GDGKVFIYHGSSLGVVVKPSQVLEGEAVGIKSFGYSLSGGLDVDGNHYPDLLVGSLAD
                TAALFRARPVLHVSQEIFIDPRAIDLEQPNCADGRLVCVDIKICFSYVAVPSSYSPSV
                ALDYMLDGDTDRRLRGQVPRVTFLSRGLDDLRHQSSGTVWLKHQHDRVCGDTVFQLQE
                NVKDKLRAIVVTLSYGLRTPPLGRQAPGQELPTVAPILNAHQPSTQRTEIHFLKQGCG
                QDKICQSNLQLERYQFCSRISDTEFQALPMDLDGRTALFALSGQPFIGLELTVTNLPS
                DPSRPQADGDDAHEAQLLVTLPASLRYSGVRALDSVEKPLCLSNDSASHVECELGNPM
                KRGAQVTFYLILSTSGITIETTELEVKLLLATISEQELDPVSVRAHVFIELPLSISGV
                ATPQQLFFSGEVKGESAMRSERDVGSKVKYEVTVSNQGQSLNTLGSAFLNIMWPHEIA
                NGKWLLYPMRVELEGGQGPGKRGICSPRPNILQLDVDSRDRRRRELGQPEPQEPPEKV
                EPSTSWWPVSSAEKRNVTLDCAQGTAKCVVFSCPLYSFDRAAVLHVWGRLWNSTFLEE
                YMAVKSLEVIVRANITVKSSIKNLLLRDASTVIPVMVYLDPMAVVVEGVPWWVILLAV
                LAGLLVLALLVLLLWKLGFFKRAKHPEATVPQYHAVKIPREDRQQFKEEKTGTIQRSN
                WGNSQWEGSDAHPILAADWHPELGPDGHPVPATA"
sig_peptide     573. .671
                /gene="ITGA7"
CDS             join(573. .778,Y12383.1:79. .206,Y12383.1:307. .386,
                Y12384.1:87. .342,Y12385.1:14. .145,Y12385.1:391. .510,
                Y12386.1:113. .320,Y12386.1:423. .616,Y12386.1:859. .947,
                Y12386.1:1057. .1184,Y12386.1:1311. .1406,
                Y12386.1:1547. .1608,Y12386.1:1710. .1879,Y12387.1:110. .259,
                Y12388.1:110. .225,Y12388.1:622. .814,Y12389.1:109. .189,
                Y12389.1:276. .355,Y12389.1:454. .528,Y12389.1:667. .769,
                Y12389.1:1256. .1432,Y12389.1:1597. .1725,Y12390.1:64. .177,
                Y12390.1:381. .479,Y12390.1:567. .692,Y12381.1:106. .218,
                Y12382.1:53. .113)
                /gene="ITGA7"
                /function="laminin receptor"
                /note="A variant"
                /codon_start=1
                /label=ITGCDSA
                /product="integrin alpha 7"
                /protein_id="CAA73024.1"
                /db_xref="GI:3378243"
                /db_xref="SPTREMBL:O88731"
                /translation="MARIPRCDFLRPPGIYYLITSLLAGLFLPPAIAFNLDVMGAIRK
                EGEPGSLFGFSVALHRQLQPRPQSWLLVGAPQALALPGQQANRTGGLFACPLSLEETD
                CYRVDIDRGANVQKESKENQWLGVSVRSQGPGGKIVTCAHRYESRQRVDQALETRDVI
                GRCFVLSQDLAIRDELDGGEWKFCEGRPQGHEQFGFCQQGTAATFSPDSHYLVFGAPG
                TYNWKGTARVELCAQGSPDLAHLDDGPYEAGGEKEQDPRLIPVPANSYLGLLFVTNID
                SSDPDQLVYKTLDPADRLTGPAGDLTLNSYLGFSIDSGKGLMRSEELSFVAGAPRANH
                KGAVVILRKDSASRLIPEVVLSGERLTSGFGYSLAVTDLNNDGWADLIVGAPYFFERQ
                EELGGAVYVYMNQGGHWADISPLRICGSPDSMFGISLAVLGDLNQDGFPDIAVGAPFD
                GDGKVFIYHGSSLGVVVKPSQVLEGEAVGIKSFGYSLSGGLDVDGNHYPDLLVGSLAD
                TAALFRARPVLHVSQEIFIDPRAIDLEQPNCADGRLVCVDIKICFSYVAVPSSYSPSV
                ALDYMLDGDTDRRLRGQVPRVTFLSRGLDDLRHQSSGTVWLKHQHDRVCGDTVFQLQE
                NVKDKLRAIVVTLSYGLRTPPLGRQAPGQELPTVAPILNAHQPSTQRTEIHFLKQGCG
                QDKICQSNLQLERYQFCSRISDTEFQALPMDLDGRTALFALSGQPFIGLELTVTNLPS
                DPSRPQADGDDAHEAQLLVTLPASLRYSGVRALDSVEKPLCLSNDSASHVECELGNPM
                KRGAQVTFYLILSTSGITIETTELEVKLLLATISEQELDPVSVRAHVFIELPLSISGV
                ATPQQLFFSGEVKGESAMRSERDVGSKVKYEVTVSNQGQSLNTLGSAFLNIMWPHEIA
                NGKWLLYPMRVELEGGQGPGKRGICSPRPNILQLDVDSRDRRRRELGQPEPQEPPEKV
                EPSTSWWPVSSAEKRNVTLDCAQGTAKCVVFSCPLYSFDRAAVLHVWGRLWNSTFLEE
                YMAVKSLEVIVRANITVKSSIKNLLLRDASTVIPVMVYLDPMAVVVEGVPWWVILLAV
                LAGLLVLALLVLLLWKCGFFRRNSPSSSFPTNYHRAHLAVQPSAMEAGGPGTVGWDSS
                SGRSTPRPPCPSTTQ"
BASE COUNT      170 a    258 c    276 g    211 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:     10
              Ratio:  4.667              Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-27 x MMY12380   ..

 Align seg 1/1  to: MMY12380  from: 1  to: 915

       2 GluAlaGluProGluAlaGluProIleMet 11
         ||||||||||||||||   ||||||::::::
     349 GAGGCGGAGCCGGAGCCGGAGCCGGTGCTC 378

seq_name: gb_in:AMAPID73

seq_documentation_block:
LOCUS       AMAPID73     1010 bp    mRNA            INV       01-SEP-1999
DEFINITION  Apis mellifera Apid73 mRNA.
ACCESSION   X72577
VERSION     X72577.1  GI:297066
KEYWORDS    antibiotic peptide; apid73 gene; Apidaecin; precursor protein.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
            Aculeata; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 1010)
  AUTHORS   Casteels-Josson,K., Capaci,T., Casteels,P. and Tempst,P.
  TITLE     Apidaecin multipeptide precursor structure: a putative mechanism
            for amplification of the insect antibacterial response
  JOURNAL   EMBO J. 12 (4), 1569-1578 (1993)
  MEDLINE   93223697
REFERENCE   2  (bases 1 to 1010)
  AUTHORS   Casteels-Josson,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1993) K. Casteels-Josson, Memorial Sloan
            Kettering Cancer Center, 1275 York Avenue, New York 10021, NY, USA
FEATURES             Location/Qualifiers
     source          1. .1010
                     /organism="Apis mellifera"
                     /db_xref="taxon:7460"
     primer_bind     1. .14
     sig_peptide     <15. .68
                     /gene="Apid73"
                     /product="secretion signal"
     CDS             <15. .866
                     /gene="Apid73"
                     /codon_start=1
                     /product="Apidaecin precursor"
                     /protein_id="CAA51169.1"
                     /db_xref="GI:4539289"
                     /db_xref="SWISS-PROT:Q06602"
                     /translation="KNFALAILVVTFVVAVFGNTNLDPPTRPTRLRREAKPEAEPGNN
                     RPVYIPQPRPPHPRLRREAEPEAEPGNNRPVYIPQPRPPHPRLRREAELEAEPGNNRP
```

```
                     VYISQPRPPHPRLRREAEPEAEPGNNRPVYIPQPRPPHPRLRREAELEAEPGNNRPVY
                     ISQPRPPHPRLRREAEPEAEPGNNRPVYIPQPRPPHPRLRREAEPEAEPGNNRPVYIP
                     QPRPPHPRLRREAEPEAEPGNNRPVYIPQPRPPHPRLRREAKPEAKPGNNRPVYIPQP
                     RPPHPRI"
     gene            15. .866
                     /gene="Apid73"
     mat_peptide     66. .863
                     /gene="Apid73"
                     /note="promultipeptide"
                     /product="Apidaecin"
     misc_signal     866. .870
                     /note="ATTTA"
     misc_signal     902. .906
                     /note="ATTTA"
     polyA_signal    906. .911
     polyA_signal    962. .967
     primer_bind     1005. .1010
BASE COUNT      319 a    279 c    185 g    227 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:      8
              Ratio:  5.250              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-27 x AMAPID73   ..

 Align seg 1/1  to: AMAPID73  from: 1  to: 1010

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||||||||||||||
     198 GAAGCTGAACCGGAAGCTGAACCC 221

seq_name: gb_in:DDU66367

seq_documentation_block:
LOCUS       DDU66367     1390 bp    mRNA            INV       28-AUG-1996
DEFINITION  Dictyostelium discoideum SapA (sapA) mRNA, partial cds.
ACCESSION   U66367
VERSION     U66367.1  GI:1513233
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 1390)
  AUTHORS   Loomis,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1996) Dept. of Biology 0322, University of
            California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-032,
            USA
FEATURES             Location/Qualifiers
     source          1. .1390
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="4"
                     /map="4.053"
                     /cell_line="AX4"
     gene            1. .1372
                     /gene="sapA"
     CDS             <1. .1372
                     /gene="sapA"
                     /note="similar to human proactivator polypeptide precursor
                     saposin A region, Swiss-Prot Accession Number P07602"
                     /codon_start=2
                     /product="SapA"
                     /protein_id="AAB06759.1"
                     /db_xref="GI:1513234"
                     /translation="LPKEYAQDCVEIINNYGVLMVRLLINRESPENVCLMMELCSKSS
                     SELIMPINVVAVGDEVDNVEGSGECLVCEFISEKIVTYLEANQTETQILQYLDNDCKL
                     LKSDNWISTCQNLIQEYEPQIIAVVEGFMAPSELCQKIKFCSSSSSTNDFDFIGSSTT
                     DCEICTFISGYAENFLEENKTLEDIIKVVDDFCKILPAAYKTDCVAMASNYIPAIIKM
                     LENDNSPGQVCQKLNLCPAPTPTSTPSTIKIDVNINKCEICETVIGFAEKYLGSALDS
                     QTVKTFLQNECNKLPGVVGDVCNEVVNEYLEFAVTQLEAKFNPTEICQLFSFCTAGSE
                     SSDASEIKINNNSYIDCQICEVLANFVVYDKNYTTHFEDKLNKLCSCLSRSLIEPCLN
                     FVSQYGSDLEQIIVQGGESAAEACNQIAQCPTSSSSSYSGSASGASSGSASGSASGSG
                     SGGYLF"
BASE COUNT      485 a    193 c    225 g    487 t
ORIGIN


alignment_scores:
            Quality:   42.00            Length:      8
              Ratio:  5.250              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-27 x DDU66367/rev   ..

 Align seg 1/1  to reverse of: DDU66367  from: 1  to: 1390

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||||||||||||||
    1341 GAGGCAGAACCTGAAGCAGAACCT 1318

seq_name: gb_in:DDU72746

seq_documentation_block:
LOCUS       DDU72746     1486 bp    mRNA            INV       19-MAR-1997
DEFINITION  Dictyostelium discoideum cysteine proteinase (cprG) mRNA, complete
            cds.
ACCESSION   U72746
VERSION     U72746.1  GI:1644501
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 1486)
  AUTHORS   Ord,T., Adessi,C., Wang,L. and Freeze,H.H.
  TITLE     The cysteine proteinase gene cprG in Dictyostelium discoideum has a
            serine-rich domain that contains GlcNAc-1-P
  JOURNAL   Arch. Biochem. Biophys. 339 (1), 64-72 (1997)
  MEDLINE   97223364
REFERENCE   2  (bases 1 to 1486)
  AUTHORS   Ord,T., Adessi,C., Wang,L. and Freeze,H.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1996) The Burnham Institute, 10901 N. Torrey
            Pines Rd. La Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .1486
                     /organism="Dictyostelium discoideum"
                     /strain="AX-2"
                     /db_xref="taxon:44689"
     gene            9. .1391
                     /gene="cprG"
     CDS             9. .1391
                     /gene="cprG"
                     /codon_start=1
                     /product="cysteine proteinase"
                     /protein_id="AAC47482.1"
                     /db_xref="GI:1644502"
                     /translation="MKVLSALCVLLVSVATAKQQLSEVEYRNAFTNWMIAHQRHYSSE
                     EFNGRYNIFKANMDYVNEWNTKGSETVLGLNVFADISNEEYRATYLGTPFDASSLEMT
                     ESDKIFDASAQVDWRTQGAVTPIKNQGQCGGCWSFSTTGATEGAQYLANGKKNLVSLS
                     EQNLIDCSGSYGNNGCEGGLMTLAFEYIINNKGIDTESSYPYTAEDGKKCKFNPKNVA
                     AQLSSYVNVTSGSESDLAAKVTQGPTSVAIDASNQSFQLYVSGIYNEPACSSTQLDHG
                     VLAVGFGTGSGSSGSHGGSQSQSAGSDSQSAGSESSQSESGSQSQSESGSQSQSQSGS
                     QSFSGSLYSGSYSGSQSGSQSGNSGAAVKQTGAGSGSGSGSGSGSGSGSVSGSASG
                     SASGSASGSSSGSNSNGGVYPTAGDYWIVKNSWGTSWGMDGYILMTKGNNNQCGIATM
                     ASRPTAVASL"
BASE COUNT      448 a    314 c    278 g    446 t
ORIGIN
```

```
alignment_scores:
            Quality:   42.00          Length:      8
              Ratio:  5.250            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-27 x DDU72746/rev

 Align seg 1/1  to reverse of: DDU72746  from: 1  to: 1486

       2 GluAlaGluProGluAlaGluPro 9
         ||||||||||||||||||||||||
    1204 GAGGCGGAACCTGAGGCGGAACCT 1181
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:38:37 ; Search time 24.75 Seconds
(without alignments)
10.001 Million cell updates/sec

Title: US-09-444-281-27
Perfect score: 59
Sequence: 1 HEAEPEAEPIM 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	66.1	486	2	US-08-942-423-3	Sequence 3, Appli
2	38	64.4	20	1	US-08-208-181A-8	Sequence 8, Appli
3	38	64.4	20	1	US-08-208-181A-10	Sequence 10, Appl
4	38	64.4	27	1	US-07-603-782A-1	Sequence 1, Appli
5	38	64.4	229	2	US-08-394-600B-20	Sequence 20, Appl
6	38	64.4	229	4	US-08-944-483-30	Sequence 30, Appl
7	38	64.4	229	5	PCT-US95-02513-20	Sequence 20, Appl
8	38	64.4	256	2	US-08-230-428B-4	Sequence 4, Appli
9	37	62.7	226	4	US-08-944-483-41	Sequence 41, Appl
10	37	62.7	227	4	US-08-944-483-40	Sequence 40, Appl
11	37	62.7	247	2	US-08-851-974-1	Sequence 1, Appli
12	37	62.7	247	2	US-08-851-974-4	Sequence 4, Appli
13	37	62.7	247	2	US-09-213-390-1	Sequence 1, Appli
14	37	62.7	247	2	US-09-213-390-4	Sequence 4, Appli
15	37	62.7	248	2	US-08-851-974-3	Sequence 3, Appli
16	37	62.7	248	2	US-09-213-390-3	Sequence 3, Appli
17	37	62.7	302	4	US-09-457-046B-18	Sequence 18, Appl
18	37	62.7	625	3	US-08-996-139-15	Sequence 15, Appl
19	37	62.7	625	4	US-08-995-659-15	Sequence 15, Appl
20	37	62.7	625	4	US-09-215-649A-15	Sequence 15, Appl
21	36	61.0	364	2	US-08-651-940-2	Sequence 2, Appli
22	36	61.0	364	4	US-09-025-580-37	Sequence 37, Appl
23	36	61.0	372	4	US-09-286-904-24	Sequence 24, Appl
24	35	59.3	36	2	US-08-942-423-12	Sequence 12, Appl
25	35	59.3	433	1	US-08-522-166-7	Sequence 7, Appli
26	35	59.3	433	1	US-08-488-382A-7	Sequence 7, Appli
27	35	59.3	433	2	US-08-480-912-7	Sequence 7, Appli
28	35	59.3	466	2	US-08-836-791-9	Sequence 9, Appli
29	35	59.3	486	2	US-08-942-423-2	Sequence 2, Appli
30	35	59.3	2713	5	PCT-US96-01735-1	Sequence 1, Appli
31	34	57.6	10	1	US-08-468-674B-86	Sequence 86, Appl
32	34	57.6	10	1	US-08-468-674B-88	Sequence 88, Appl
33	34	57.6	10	1	US-08-780-571-86	Sequence 86, Appl
34	34	57.6	10	1	US-08-780-571-88	Sequence 88, Appl
35	34	57.6	10	4	US-09-012-669F-45	Sequence 45, Appl
36	34	57.6	13	4	US-08-932-082-6	Sequence 6, Appli
37	34	57.6	59	1	US-08-306-871-25	Sequence 25, Appl
38	34	57.6	59	1	US-08-569-959-25	Sequence 25, Appl
39	34	57.6	65	1	US-08-468-674B-71	Sequence 71, Appl
40	34	57.6	65	1	US-08-780-571-71	Sequence 71, Appl
41	34	57.6	88	1	US-08-468-674B-75	Sequence 75, Appl
42	34	57.6	88	1	US-08-780-571-75	Sequence 75, Appl
43	34	57.6	326	4	US-09-058-389A-3	Sequence 3, Appli
44	34	57.6	347	3	US-09-059-369-2	Sequence 2, Appli
45	34	57.6	456	4	US-09-058-389A-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-942-423-3
; Sequence 3, Application US/08942423
; Patent No. 5891673
;  GENERAL INFORMATION:
;    APPLICANT:  Hashimoto, Yasuhiro
;    APPLICANT:  Takemoto, Yoshihiro
;    TITLE OF INVENTION:  Lck Binding Protein
;    NUMBER OF SEQUENCES:  68
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Syntex (U.S.A.) Inc.
;      STREET:  3401 Hillview Ave.
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/942,423
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,715
;      FILING DATE:  23-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Peries, Rohan
;      REGISTRATION NUMBER:  35,752
;      REFERENCE/DOCKET NUMBER:  28260
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 852-1698
;      TELEFAX:  (415) 496-3529
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  LCK BINDING PROTEIN
US-08-942-423-3
```

```
  Query Match             66.1%;  Score 39;  DB 2;  Length 486;
  Best Local Similarity   77.8%;  Pred. No. 35;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEP 9
          :|||||  ||
Db    360 YEAEPEPEP 368


RESULT   2
US-08-208-181A-8
; Sequence 8, Application US/08208181A
; Patent No. 5654167
;  GENERAL INFORMATION:
;    APPLICANT:  Gabay, Joelle E.
;    APPLICANT:  Nathan, Carl F.
;    TITLE OF INVENTION:  ANTIMICROBIAL PROTEINS, COMPOSITIONS
;    TITLE OF INVENTION:  CONTAINING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  2200 Sand Hill Road, Suite 100
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/208,181A
;      FILING DATE:  08-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bozicevic, Karl
;      REGISTRATION NUMBER:  28,807
;      REFERENCE/DOCKET NUMBER:  06514/024002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 322-5070
;      TELEFAX:  (415) 854-0875
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  N-terminal
US-08-208-181A-8


  Query Match             64.4%;  Score 38;  DB 1;  Length 20;
  Best Local Similarity   54.5%;  Pred. No. 1.6;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db      5 HEAQPHSRPYM 15


RESULT   3
US-08-208-181A-10
; Sequence 10, Application US/08208181A
; Patent No. 5654167
;  GENERAL INFORMATION:
;    APPLICANT:  Gabay, Joelle E.
;    APPLICANT:  Nathan, Carl F.
;    TITLE OF INVENTION:  ANTIMICROBIAL PROTEINS, COMPOSITIONS
;    TITLE OF INVENTION:  CONTAINING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  2200 Sand Hill Road, Suite 100
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/208,181A
;      FILING DATE:  08-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bozicevic, Karl
;      REGISTRATION NUMBER:  28,807
;      REFERENCE/DOCKET NUMBER:  06514/024002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 322-5070
;      TELEFAX:  (415) 854-0875
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  N-terminal
US-08-208-181A-10


  Query Match             64.4%;  Score 38;  DB 1;  Length 20;
  Best Local Similarity   54.5%;  Pred. No. 1.6;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db      5 HEAQPHSRPYM 15


RESULT   4
US-07-603-782A-1
; Sequence 1, Application US/07603782A
; Patent No. 5200319
;  GENERAL INFORMATION:
;    APPLICANT:  Arnaout, Amin M.
;    APPLICANT:  McCluskey, Robert T.
;    APPLICANT:  Niles, John L.
;    TITLE OF INVENTION:  Diagnosis of Glomerulonephritis
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/603,782A
;      FILING DATE:  19901025
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:    1
;      APPLICATION NUMBER:  07/428,286
;      FILING DATE:  27-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/034002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  27
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
US-07-603-782A-1


  Query Match             64.4%;  Score 38;  DB 1;  Length 27;
  Best Local Similarity   54.5%;  Pred. No. 2.3;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db      5 HEAQPHSRPYM 15


RESULT   5
US-08-394-600B-20
; Sequence 20, Application US/08394600B
; Patent No. 5843693
;  GENERAL INFORMATION:
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McAndrews, Held & Malloy, Ltd.
;      STREET:  500 West Madison Street; 34th Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60661
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/394,60QB
;      FILING DATE:  02/27/95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donald J. Pochopien
;      REGISTRATION NUMBER:  32,167
;      REFERENCE/DOCKET NUMBER:  820.005/11850US05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/707-8889
;      TELEFAX:  312/707-9155
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  229 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-394-600B-20


  Query Match             64.4%;  Score 38;  DB 2;  Length 229;
  Best Local Similarity   54.5%;  Pred. No. 23;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db      5 HEAQPHSRPYM 15


RESULT   6
US-08-944-483-30
; Sequence 30, Application US/08944483
; Patent No. 6232456
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STEWART, KENT D.
;    APPLICANT:  STROUPE, STEVEN D.
;    TITLE OF INVENTION:  NOVEL SERINE PROTEASE REAGENTS
;    TITLE OF INVENTION:  AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
;    TITLE OF INVENTION:  OF THE PROSTATE
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,483
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6183.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  229 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6232456e
US-08-944-483-30


  Query Match             64.4%;  Score 38;  DB 4;  Length 229;
  Best Local Similarity   54.5%;  Pred. No. 23;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 HEAEPEAEPIM 11
          |||:| : | |
Db      5 HEAQPHSRPYM 15


RESULT   7
PCT-US95-02513-20
; Sequence 20, Application PC/TUS9502513
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    TITLE OF INVENTION:  Compositions for the Inhibition of
;    TITLE OF INVENTION:  Protein Hormone Formation and Uses Thereof
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02513
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams Jr., Joseph A.
;      REGISTRATION NUMBER:  38,659
;      REFERENCE/DOCKET NUMBER:  27527/32404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6900
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  229 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-02513-20


  Query Match             64.4%;  Score 38;  DB 5;  Length 229;
  Best Local Similarity   54.5%;  Pred. No. 23;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
             |||:| : | |
Db       5 HEAQPHSRPYM 15


RESULT   8
US-08-230-428B-4
; Sequence 4, Application US/08230428B
; Patent No. 5998378
;  GENERAL INFORMATION:
;    APPLICANT:  Kriegler, Michael
;    APPLICANT:  Perez, Carl
;    APPLICANT:  Halenbeck, Robert F.
;    APPLICANT:  Jewell, David A.
;    APPLICANT:  Koths, Kirston E.
;    TITLE OF INVENTION:  Compositions For The Inhibition Of TNF
;    TITLE OF INVENTION:  Hormone Formation And Uses Thereof (As Amended)
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION  Intellectual Property - R440
;      STREET:  4560 Horton Street, P.O. Box 8097
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,428B
;      FILING DATE:  19-APR-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/905,546
;      FILING DATE:  25-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/395,253
;      FILING DATE:  16-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savereide, Paul B.
;      REGISTRATION NUMBER:  36,914
;      REFERENCE/DOCKET NUMBER:  0820.004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2718
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  256 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-230-428B-4


  Query Match             64.4%;  Score 38;  DB 2;  Length 256;
  Best Local Similarity   54.5%;  Pred. No. 26;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
             |||:| : | |
Db      32 HEAQPHSRPYM 42


RESULT   9
US-08-944-483-41
; Sequence 41, Application US/08944483
; Patent No. 6232456
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STEWART, KENT D.
;    APPLICANT:  STROUPE, STEVEN D.
;    TITLE OF INVENTION:  NOVEL SERINE PROTEASE REAGENTS
;    TITLE OF INVENTION:  AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
;    TITLE OF INVENTION:  OF THE PROSTATE
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,483
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
```

```
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6183.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  226 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6232456e
US-08-944-483-41


  Query Match             62.7%;  Score 37;  DB 4;  Length 226;
  Best Local Similarity   54.5%;  Pred. No. 33;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db       5 HEAKPHSRPYM 15


RESULT  10
US-08-944-483-40
; Sequence 40, Application US/08944483
; Patent No. 6232456
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STEWART, KENT D.
;    APPLICANT:  STROUPE, STEVEN D.
;    TITLE OF INVENTION:  NOVEL SERINE PROTEASE REAGENTS
;    TITLE OF INVENTION:  AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
;    TITLE OF INVENTION:  OF THE PROSTATE
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,483
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6183.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  227 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6232456e
US-08-944-483-40


  Query Match             62.7%;  Score 37;  DB 4;  Length 227;
  Best Local Similarity   54.5%;  Pred. No. 33;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db       5 HEAKPHSRPYM 15


RESULT  11
US-08-851-974-1
; Sequence 1, Application US/08851974
; Patent No. 5858758
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN SERINE PROTEASE PRECURSOR
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/851,974
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0288 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  247 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  NGANNOT01
;      CLONE:  854243
US-08-851-974-1


  Query Match             62.7%;  Score 37;  DB 2;  Length 247;
  Best Local Similarity   54.5%;  Pred. No. 36;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      25 HEAKPHSRPYM 35
```

```
RESULT 12
US-08-851-974-4
; Sequence 4, Application US/08851974
; Patent No. 5858758
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN SERINE PROTEASE PRECURSOR
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/851,974
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0288 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  247 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  306682
US-08-851-974-4


  Query Match             62.7%;  Score 37;  DB 2;  Length 247;
  Best Local Similarity   54.5%;  Pred. No. 36;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
             |||:| : | |
Db      25 HEAKPHSRPYM 35


RESULT 13
US-09-213-390-1
; Sequence 1, Application US/09213390
; Patent No. 5965711
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN SERINE PROTEASE PRECURSOR
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/213,390
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/851,974
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0288 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  247 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  NGANNOT01
;      CLONE:  854243
US-09-213-390-1


  Query Match             62.7%;  Score 37;  DB 2;  Length 247;
  Best Local Similarity   54.5%;  Pred. No. 36;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
             |||:| : | |
Db      25 HEAKPHSRPYM 35


RESULT 14
US-09-213-390-4
; Sequence 4, Application US/09213390
; Patent No. 5965711
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN SERINE PROTEASE PRECURSOR
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/213,390
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/851,974
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0288 US
;    TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 306682
US-09-213-390-4

Query Match 62.7%; Score 37; DB 2; Length 247;
Best Local Similarity 54.5%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 HEAEPEAEPIM 11
|||:| : | |
Db 25 HEAKPHSRPYM 35

RESULT 15
US-08-851-974-3
; Sequence 3, Application US/08851974
; Patent No. 5858758
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN SERINE PROTEASE PRECURSOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,974
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0288 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 206690
US-08-851-974-3

Query Match 62.7%; Score 37; DB 2; Length 248;
Best Local Similarity 54.5%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 HEAEPEAEPIM 11
|||:| : | |
Db 25 HEAKPHSRPYM 35

Search completed: January 4, 2002, 08:40:57
Job time: 140 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:38:37 ; Search time 53.46 Seconds
(without alignments)
15.241 Million cell updates/sec

Title: US-09-444-281-27
Perfect score: 59
Sequence: 1 HEAEPEAEPIM 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	66.1	6797	22	AAB31558	Pimaricin biosynth
2	38	64.4	27	12	AAR11996	N-terminal of p29
3	38	64.4	27	14	AAR34444	N-terminus of p29.
4	38	64.4	237	13	AAR20509	Human proteinase 3
5	38	64.4	256	16	AAR45403	Deduced sequence o
6	38	64.4	256	16	AAR85639	MY17 preproPR-3.
7	37	62.7	149	22	AAU00451	Protein encoded by
8	37	62.7	173	22	AAB72305	Sunflower glutamic
9	37	62.7	246	12	AAR13253	Human Cytotoxic Ce
10	37	62.7	247	20	AAW84158	A human serine pro
11	37	62.7	261	19	AAW54091	Homo sapiens BE2 s
12	37	62.7	267	21	AAY90291	Human peptidase, H
13	37	62.7	267	22	AAB20156	Human protein SECP
14	37	62.7	278	21	AAB58142	Lung cancer associ
15	37	62.7	302	22	AAE00235	Protein encoded by
16	37	62.7	553	19	AAW63703	Truncated rat rSK3
17	37	62.7	567	20	AAW88788	Polypeptide fragme
18	37	62.7	568	21	AAY85658	Human Acinus S pro
19	37	62.7	571	20	AAY41704	Human PRO351 prote
20	37	62.7	571	21	AAB44260	Human PRO351 (UNQ3
21	37	62.7	571	21	AAB24046	Human PRO351 prote
22	37	62.7	583	21	AAY85659	Human Acinus S' pr
23	37	62.7	625	19	AAW83200	Murine osteoclast
24	37	62.7	625	19	AAW69958	Murine NF-kB recep
25	37	62.7	625	19	AAW68294	Murine NF-kB recep
26	37	62.7	625	21	AAY59509	OBM binding protei
27	37	62.7	625	21	AAY53649	A mouse receptor a
28	37	62.7	625	22	AAE04427	Murine receptor ac
29	37	62.7	625	22	AAE01994	Murine RANK (recep
30	37	62.7	732	19	AAW63715	Rat rSK3 protein.
31	37	62.7	1341	21	AAY85657	Human Acinus L pro
32	36	61.0	75	21	AAG02199	Human secreted pro
33	36	61.0	108	21	AAB33906	Human secreted pro
34	36	61.0	161	21	AAG42933	Arabidopsis thalia
35	36	61.0	181	21	AAG17997	Arabidopsis thalia
36	36	61.0	181	21	AAG47650	Arabidopsis thalia
37	36	61.0	183	21	AAB41769	Human ORFX ORF1533
38	36	61.0	192	22	AAU14154	Human novel protei
39	36	61.0	196	21	AAG42932	Arabidopsis thalia
40	36	61.0	213	21	AAB43275	Human ORFX ORF3039
41	36	61.0	216	21	AAG17996	Arabidopsis thalia
42	36	61.0	216	21	AAG47649	Arabidopsis thalia
43	36	61.0	252	21	AAG15889	Arabidopsis thalia
44	36	61.0	254	21	AAG47624	Arabidopsis thalia
45	36	61.0	264	21	AAG47623	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB31558
ID AAB31558 standard; Protein; 6797 AA.
XX
AC AAB31558;
XX
DT 20-APR-2001 (first entry)
XX
DE Pimaricin biosynthesis associated polyketide synthase polypeptide.
XX
KW Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW anticancer; pimaricin.
XX
OS Streptomyces natalensis.
XX
PN WO200077222-A1.
XX
PD 21-DEC-2000.
XX
PF 14-JUN-2000; 2000WO-EP06227.
XX
PR 14-JUN-1999; 99EP-0201893.
XX
PA (STAM) DSM NV.
XX
PI Martin JF, Aparicio JF, Colina AJ;
XX
DR WPI; 2001-080693/09.
DR N-PSDB; AAF24892.
XX
PT New polynucleotides encoding enzymes involved in the biosynthesis of
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT the synthesis of new compounds -
XX

```
PS   Disclosure; Page 81-101; 116pp; English.
XX
CC   The present sequence represents a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence   6797 AA;


  Query Match             66.1%;  Score 39;  DB 22;  Length 6797;
  Best Local Similarity   77.8%;  Pred. No. 1.5e+03;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEP 9
           |||||  ||
Db     463 heaepapep 471


RESULT   2
AAR11996
ID   AAR11996 standard; peptide; 27 AA.
XX
AC   AAR11996;
XX
DT   26-JUL-1991  (first entry)
XX
DE   N-terminal of p29 protein.
XX
KW   Wegener's granulomatosis; monoclonal antibodies; autoantibodies;
KW   glomerulonephritis; serine protease; antigen.
XX
OS   Homo sapiens.
XX
PN   WO9106572-A.
XX
PD   16-MAY-1991.
XX
PF   29-OCT-1990;   90WO-US06277.
XX
PR   27-OCT-1989;   89US-0428286.
XX
PA   (GEHO-) GEN HOSPITAL CORP.
XX
PI   Arnaout M,  Mccluskey RT,  Niles J;
XX
DR   WPI; 1991-164137/22.
XX
PT   Purified P29 protein - used to detect auto-antibodies diagnostic
PT   for Wegener's granulomatosis and conditions associated with
PT   glomerulo nephritis
XX
PS   Claim 1; Page 22; 33pp; English.
XX
CC   The p29 protein is a 29 kD antigen which was prepd. by affinity
CC   purificn. from neutrophil acid extract, using 1E8 monoclonal
CC   antibodies.   The purified antigen migrated on SDS-PAGE as three
CC   close bands, with major component at 29 kD under non-reducing
CC   conditions.  It reacted with autoantibodies from patients sera
CC   indicating identity between Wegener's granulomatois autoantigen.
CC   On isofocusing gels it had a pI of 9.2-9.4.  It is a novel serine
CC   protease showing homology with leukocyte elastase and cathepsin G.
CC   The protein or monoclonal Abs can be used to test for the presence
CC   of autoantibodies diagnostic for Wegener's granulomatosis. In
CC   combination with myeloperoxidase (and/or Abs against it), p29 MAbs
CC   can also be used to test for autoantibodies associated with pauci
CC   immune necrotizing and/or crescentric glomerulonephritis.
XX
SQ   Sequence   27 AA;


  Query Match             64.4%;  Score 38;  DB 12;  Length 27;
  Best Local Similarity   54.5%;  Pred. No. 7.7;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db       5 heaqphsrpym 15


RESULT   3
AAR34444
ID   AAR34444 standard; peptide; 27 AA.
XX
AC   AAR34444;
XX
DT   29-JUL-1993  (first entry)
XX
DE   N-terminus of p29.
XX
KW   Pauci-immune necrotising; crescentic glomerulonephritis; neutrophil;
KW   autoantibodies; Wegener's granulomatosis; microscopic polyarteritis
KW   nodosa; primary NCGN.
XX
OS   Homo sapiens.
XX
PN   US5200319-A.
XX
PD   06-APR-1993.
XX
PF   27-OCT-1989;   89US-0428286.
XX
PR   27-OCT-1989;   89US-0428286.
PR   25-OCT-1990;   90US-0603782.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Arnaout MA,  Mccluskey RT,  Niles JL;
XX
DR   WPI; 1993-133732/16.
XX
PT   Diagnosis of pauci-immune necrotising and/or crescentic
PT   glomerulonephritis - using p29 protein isolated from
PT   neutrophil(s), and which binds auto-antibodies present in serum
PT   of patients with Wegener's granulomatosis
XX
PS   Disclosure; Page 12; 13pp; English.
XX
CC   The p29 protein can be solated from neutrophils, has a mol. wt. of
CC   29 kD (SDS PAGE), binds diisopropyl fluorophosphate, has a pI of 9.2-
CC   9.4, is capable of binding auto antibodies present in the sera of
CC   individuals afflicted with Wegener's granulomatosis and has the N-
CC   terminal sequence shown.  The p29 protein can be used for diagnosis
CC   of Wegener's granulomatyosis, microscopic polyarteritis nodosa and
CC   primary NCGN.
XX
SQ   Sequence   27 AA;


  Query Match             64.4%;  Score 38;  DB 14;  Length 27;
  Best Local Similarity   54.5%;  Pred. No. 7.7;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db       5 heaqphsrpym 15


RESULT   4
AAR20509
ID   AAR20509 standard; Protein; 237 AA.
```

```
XX
AC   AAR20509;
XX
DT   06-MAY-1992  (first entry)
XX
DE   Human proteinase 3.
XX
KW   Serine protease; polymerase chain reaction; PCR; autoantibody.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..6
FT                   /label= signal
FT                   /note= "part of signal peptide"
FT   Peptide         7..8
FT                   /label= propeptide
FT   Protein         9..237
FT                   /label= proteinase_3
XX
PN   WO9200378-A.
XX
PD   09-JAN-1992.
XX
PF   20-JUN-1991;   91WO-1001142.
XX
PR   22-JUN-1990;   90DE-4019984.
XX
PA   (GBFB-) GBF GES BIOTECHN FO.
XX
PI   Jenne D,  Tschopp J,  Ludemann J,  Utecht B,  Gross LB;
XX
DR   WPI; 1992-041560/05.
DR   N-PSDB; AAQ20727.
XX
PT   DNA sequence coding for proteinase 3 - allows mass prodn. of
PT   enzyme, with modifications for improved therapy and diagnosis of
PT   Wegener's granulomatosis
XX
PS   Disclosure; Fig 3; 21pp; German.
XX
CC   The cDNA sequence corrresponding to the coding region of the
CC   proteinase 3 gene was obtained by PCR amplification of cDNA from
CC   cell line U937 (see AAQ22231 and AAQ22232 for PCR primer sequences). An
CC   amplified product of 784bp was obtained. The product was cut with
CC   EcoRI and cloned into EcoRI-digested M13mp18/BAP. The complete
CC   sequence was determined and the amino acid sequence it encodes was
CC   deduced from it.
XX
SQ   Sequence   237 AA;

  Query Match             64.4%;  Score 38;  DB 13;  Length 237;
  Best Local Similarity   54.5%;  Pred. No. 71;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      13 heaqphsrpym 23


RESULT   5
AAR45403
ID   AAR45403 standard; Protein; 256 AA.
XX
AC   AAR45403;
XX
DT   18-JUL-1994  (first entry)
XX
DE   Deduced sequence of human proteinase-3 (PR-3).
XX
KW   TNF convertase; proteinase-3; PR-3; tumour necrosis factor;
KW   proTNF.
XX
OS   Homo sapiens.
XX
PN   WO9400555-A.
XX
PD   06-JAN-1994.
XX
PF   25-JUN-1993;   93WO-US06120.
XX
PR   25-JUN-1992;   92US-0905546.
XX
PA   (CETU ) CETUS ONCOLOGY CORP.
XX
PI   Halenbeck RF,  Jewell DA,  Koths KE,  Kriegler M,  Perez C;
XX
DR   WPI; 1994-026195/03.
DR   N-PSDB; AAQ54498.
XX
PT   Cpds. which inhibit formation of mature tumour necrosis factor
PT   from its precursor - identified using TNF convertase, e.g.
PT   mutein(s), antibodies or peptide phosphonate(s), for preventing
PT   and treating sepsis, AIDS, auto-immune disease etc.
XX
PS   Disclosure; Fig 2; 69pp; English.
XX
CC   ProTNF refers to TNF having a molecular weight of about 26,000,
CC   which is the prohormone form of TNFa. ProTNF is cleaved to a lower
CC   molecular weight 'mature' form, pref. 17kD, which, in its multimeric
CC   (usually trimeric) form, is substantially involved in producing life-
CC   threatening physiological changes associated with sepsis. ProTNF is
CC   cleaved by convertase. One TNF convertase is serine protease
CC   proteinase-3, also called PR-3, P-29B or myeloblasin. A suitable
CC   source of convertase is the HL60 cell line (or extracts, or the
CC   culture media in which it is grown). The convertase produced by
CC   HL60 has been sequenced and is identical to the known lymphocyte
CC   serine protease PR-3 which has other activities unrelated to TNF
CC   processing.
XX
SQ   Sequence   256 AA;

  Query Match             64.4%;  Score 38;  DB 15;  Length 256;
  Best Local Similarity   54.5%;  Pred. No. 76;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      32 heaqphsrpym 42


RESULT   6
AAR85639
ID   AAR85639 standard; Protein; 256 AA.
XX
AC   AAR85639;
XX
DT   23-APR-1996  (first entry)
XX
DE   MY17 preproPR-3.
XX
KW   PR-3; preproPR-3; MY17; human neutrophil protease-3; serine protease;
KW   tumour necrosis factor alpha; TNFalpha; HL60; MY17; B cell; T cell;
KW   tumour necrosis factor alpha convertase; cytokine; septic shock;
KW   rheumatoid arthritis; cachexia; cerebral malaria; graft-host disease;
KW   ischaemia/reperfusion injury; autoimmune disease; AIDS.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /note= "leader sequence present only in preproPR-3"
FT   Peptide         26..27
FT                   /note= "dipeptide present in proPR-3"
```

```
FT   Protein          28..256
FT                    /note= "mature PR-3"
XX
PN   WO9524501-A1.
XX
PD   14-SEP-1995.
XX
PF   02-MAR-1995;   95WO-US02513.
XX
PR   28-FEB-1995;   95US-0395456.
PR   07-MAR-1994;   94US-0208574.
PR   19-APR-1994;   94US-0230428.
PR   27-FEB-1995;   95US-0394600.
XX
PA   (CETU ) CETUS ONCOLOGY CORP.
XX
PI   Halenback RF,  Jewell DA,  Koths KE,  Kriegler M,  Perez C;
XX
DR   WPI; 1995-328287/42.
DR   N-PSDB; AAT02565.
XX
PT   Identification of inhibitors of mature TNFalpha prodn.  - useful for
PT   treatment of septic shock, rheumatoid arthritis, etc..
XX
PS   Example 2; Page 82; 96pp; English.
XX
CC   This sequence represents the preproPR-3.  PR-3 is active recombinant
CC   human neutrophil protease-3.  PR-3 is a serine protease, and is a tumour
CC   necrosis factor alpha (TNFalpha) convertase.  The cDNA encoding this
CC   sequence was isolated from the HL60 cell clone MY17.  The mature PR-3 can
CC   be used in the method of the invention for identifying agents that
CC   inhibit cleavage of proTNFalpha to mature TNFalpha.  In the method,
CC   proTNFalpha is incubated with PR-3 (or another TNFalpha convertase), and
CC   the cleavage of the proTNFalpha is measured by a colourimetric assay.
CC   This is then repeated in the presence of a test compound that is thought
CC   to inhibit this process.  The results of the two reactions are then
CC   compared to determine whether the test compound is an inhibitor.  The
CC   cleavage inhibitors can be used in the treatment of septic shock,
CC   rheumatoid arthritis, cachexia, cerebral malaria, ischaemia/reperfusion
CC   injury, graft-host disease, autoimmune diseases, and AIDS.  PR-3 can be
CC   used to treat unwanted B cell/T cell interaction by contacting it with
CC   T cells to cause the release of membrane-bound cytokines.
XX
SQ   Sequence   256 AA;


  Query Match             64.4%;  Score 38;  DB 16;  Length 256;
  Best Local Similarity   54.5%;  Pred. No. 76;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      32 heaqphsrpym 42


RESULT   7
AAU00451
ID   AAU00451 standard; Protein; 149 AA.
XX
AC   AAU00451;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Protein encoded by sugarcane promoter cDNA clones c51, c511 and c512.
XX
KW   Sugarcane promoter region; monocotyledonous plant; stem tissue;
KW   insecticide; herbicide; disease resistance; improved food content;
KW   beta-glucuronidase; GUS; starch biosynthesis; fatty acid biosynthesis;
KW   ADP-glucose pyrophosphorylase; sucrose metabolism; clone c51; c511; c512.
XX
OS   Saccharum sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 41
FT                    /note= "Encoded by ACG in cDNA clone c51 (AAS01022)"
FT   Misc-difference 92
FT                    /label= Ala, Val
FT                    /note= "Residue 92 is Ala as deduced from cDNA clones
FT                           c51 and c511 (AAS01022-AAS01023) and Val as
FT                           deduced from clone c512 (AAS01024)"
FT   Misc-difference 103
FT                    /note= "Encoded by TAG in cDNA clone c511 (AAS01023)"
FT   Misc-difference 125
FT                    /label= Val, Ala
FT                    /note= "Residue 125 is Val as deduced from cDNA clone
FT                           c51 (AAS01022) and Ala as deduced from clones
FT                           c511 and c512 (AAS01023-AAS01024)"
FT   Misc-difference 130
FT                    /note= "Encoded by TCC in cDNA clone c511 (AAS01023)"
FT   Misc-difference 135
FT                    /note= "Encoded by GCC in cDNA clone c511 (AAS01023)"
XX
PN   WO200118211-A1.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-AU01033.
XX
PR   02-SEP-1999;   99AU-0002625.
XX
PA   (UYQU ) UNIV QUEENSLAND.
XX
PI   Potier B,  Birch RG;
XX
DR   WPI; 2001-218560/22.
DR   N-PSDB; AAS01022, AAS01023, AAS01024.
XX
PT   New sugarcane plant promoters for directing expression of heterologous
PT   nucleic acids in a constitutive or tissue-specific manner in
PT   monocotyledonous plants  -
XX
PS   Claim 40; Fig 15; 107pp; English.
XX
CC   The present sequence represents the polypeptide encoded by sugarcane
CC   plant promoter cDNA isolated from clones c51, c511 and c512. Clone c51,
CC   c511 and c512 promoter cDNAs are 3 of 11 promoter regions of a
CC   transcribable DNA sequence isolated from various sugarcane cDNA clones
CC   (AAS01021-AAS01031). Also described are 4 promoter regions of specific
CC   transcribed DNA sequences (AAS01032-AAS01035). The nucleic acids are
CC   useful for producing transgenic plants, having an altered phenotype and
CC   for driving expression of a foreign or endogenous DNA sequence, which
CC   encode agronomic properties including insecticide, herbicide, disease
CC   resistance, stress tolerance and improved food content, or increased
CC   yields. The foreign or endogenous DNA sequence may comprise a region
CC   transcribed into an antisense RNA or ribozyme that modulates the
CC   expression of a corresponding target gene, or it may encode
CC   beta-glucuronidase (GUS), luciferase, neomycin phosphotransferase, a
CC   product conferring herbicide tolerance, a product affecting starch
CC   biosynthesis or modification, ADP-glucose pyrophosphorylase, a product
CC   involved in fatty acid biosynthesis, a product conferring insect
CC   resistance, a product altering sucrose metabolism or a gene encoding
CC   valuable pharmaceuticals, e.g. antibiotics, secondary metabolites or
CC   vaccines. The promoters are capable of directing high level expression in
CC   many or all cells of a plant, preferentially in stem or meristem tissue
CC   of monocotyledonous plants.
XX
SQ   Sequence   149 AA;


  Query Match             62.7%;  Score 37;  DB 22;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 AEPEAEP 9
           |||||||
Db      69 aepeaep 75
```

```
RESULT   8
AAB72305
ID   AAB72305 standard; Protein; 173 AA.
XX
AC   AAB72305;
XX
DT   16-MAY-2001  (first entry)
XX
DE   Sunflower glutamic acid rich protein (GRP) amino acid sequence.
XX
KW   Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
KW   NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
KW   pathogen resistance; abscisic acid metabolism.
XX
OS   Helianthus annuus.
XX
PN   WO200112801-A2.
XX
PD   22-FEB-2001.
XX
PF   17-AUG-2000; 2000WO-US22961.
XX
PR   18-AUG-1999;   99US-0149656.
PR   23-MAY-2000; 2000US-0206405.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
PA   (CURA-) CURAGEN CORP.
XX
PI   Bidney DL,  Crasta OR,  Hu X,  Lu G;
XX
DR   WPI; 2001-211215/21.
DR   N-PSDB; AAF77209.
XX
PT   Novel isolated defence-related signalling gene isolated from sunflower
PT   encoding neoxanthin cleavage enzyme, amino acid permease or glutamic
PT   acid-rich protein useful for increasing resistance of plant to a
PT   pathogen -
XX
PS   Claim 6; Page 107; 135pp; English.
XX
CC   This invention relates to defence-related signalling genes isolated from
CC   the sunflower (Helianthus annuus). The genes encode a neoxanthin cleavage
CC   enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
CC   protein (GRP). The signalling gene is useful for increasing the
CC   resistance of a plant to a pathogen such as fungus, virus, bacterium,
CC   nematode or insect (e.g. European corn borer), preferably
CC   Sclerotinia spp., Phoma spp., or Phomopsis spp, by stably incorporating a
CC   construct containing the gene into the genome of the plant. The gene is
CC   useful for regulating gene expression in a plant, in response to a
CC   stimulus such as infection with a pathogen, damage from a pathogen,
CC   hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
CC   oxalic acid or expression of a gene encoding oxalic acid oxidase. The
CC   genes are also useful for stem-preferred regulation of gene expression in
CC   a plant. The genes are useful in agriculture, particularly in the
CC   breeding of crop plants with improved agronomic traits, for modifying
CC   abscisic acid (ABA) metabolism and for modifying amino acid transport and
CC   content in  plants. The present sequence represents the sunflower
CC   glutamic acid rich protein (GRP) amino acid sequence.
XX
SQ   Sequence   173 AA;


  Query Match             62.7%;  Score 37;  DB 22;  Length 173;
  Best Local Similarity   87.5%;  Pred. No. 75;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEP 9
            | ||||||
Db      76 epepeaep 83
```

```
RESULT   9
AAR13253
ID   AAR13253 standard; Protein; 246 AA.
XX
AC   AAR13253;
XX
DT   11-OCT-1991  (first entry)
XX
DE   Human Cytotoxic Cell Protease-X.
XX
KW   hCCPX inhibitor; cytotoxic T-lymphocytes; CTL.
XX
OS   Homo sapiens.
XX
PN   WO9110685-A.
XX
PD   25-JUL-1991.
XX
PF   17-JAN-1991;   91WO-US00340.
XX
PR   19-JAN-1990;   90US-0467880.
XX
PA   (SERA-) SERAGEN INC.
XX
PI   Bleackley RC,  Lobe CG,  Paetkau VH,  James MN,  Murphy M;
XX
DR   WPI; 1991-237989/32.
DR   N-PSDB; AAQ12865, AAQ12866.
XX
PT   DNA vectors, and inhibitors of cytotoxic cell protease - for
PT   treatment of auto-immune diseases e.g. pernicious anaemia,
PT   rheumatoid arthritis, allo-graft rejection etc.
XX
PS   Claim 8; Fig 9; 62pp; English.
XX
CC   The hCCPX gene was isolated from cytotoxic T-cell lymphocytes using
CC   murine CCP cDNA sequences as probes. The cDNA sequence (i.e. exons
CC   only) was deduced from the genomic sequence. This is the amino acid
CC   sequence deduced from the coding regions. A structural analysis of
CC   the deduced protein was used to design competitive inhibitors of the
CC   protease.  See also AAQ12862-5 and AAR13254-R13262.
XX
SQ   Sequence   246 AA;


  Query Match             62.7%;  Score 37;  DB 12;  Length 246;
  Best Local Similarity   54.5%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
             |||:| : | |
Db      25 heakphsrpym 35
```

```
RESULT  10
AAW84158
ID   AAW84158 standard; Protein; 247 AA.
XX
AC   AAW84158;
XX
DT   25-FEB-1999  (first entry)
XX
DE   A human serine protease precursor (HSPP) protein.
XX
KW   Human serine protease precursor; HSPP; Incyte clone 854243;
KW   immunological disorder; cancer; antagonist; allergy; asthma;
KW   Crohn's disease; multiple sclerosis; rheumatoid arthritis; glioma;
KW   lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   71
```

```
FT                   /note= "potential N-linked glycosylation site"
FT   Modified-site   104
FT                   /note= "potential N-linked glycosylation site"
FT   Disulfide-bond  49
FT                   /note= "potential disulphide bridging site"
FT   Disulfide-bond  65
FT                   /note= "potential disulphide bridging site"
FT   Disulfide-bond  142
FT                   /note= "potential disulphide bridging site"
FT   Disulfide-bond  173
FT                   /note= "potential disulphide bridging site"
FT   Disulfide-bond  188
FT                   /note= "potential disulphide bridging site"
FT   Disulfide-bond  209
FT                   /note= "potential disulphide bridging site"
FT   Disulfide-bond  230
FT                   /note= "potential disulphide bridging site"
FT   Region          37..48
FT                   /note= "potential substrate binding region"
XX
PN   WO9850424-A2.
XX
PD   12-NOV-1998.
XX
PF   06-MAY-1998;   98WO-US09096.
XX
PR   07-MAY-1997;   97US-0851974.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Corley NC,  Hillman JL,  Shah P;
XX
DR   WPI; 1999-034707/03.
DR   N-PSDB; AAV82707.
XX
PT   New human serine protease precursor and related nucleic acid,
PT   vectors - for treatment, prevention and diagnosis of immunological
PT   disorders and cancer
XX
PS   Claim 1; Fig 1A-C; 56pp; English.
XX
CC   The present sequence represents a human serine protease precursor (HSPP)
CC   protein. Nucleic acids encoding HSPP were first identified in Incyte
CC   clone 854243 from the ganglioneuroma tissue cDNA library. The protein
CC   has homology with the rat natural killer cell protease-1 precursor
CC   (RNKP-1) and a human serine protease from cytotoxic T lymphocytes (SECT).
CC   As increased levels of the protein are associated with immunological
CC   disorders and cancers, antagonists are used to treat or prevent
CC   conditions such as allergy, asthma, Crohn's disease, multiple sclerosis,
CC   rheumatoid arthritis, glioma, lymphoma, myeloma etc. The HSPP nucleic
CC   acid sequence and its fragments are used as antisense/ribozyme
CC   therapeutics, for detecting and quantifying gene expression (as probes
CC   and primers in standard hybridisation and amplification assays), and for
CC   isolating related sequences and for chromosome mapping.
XX
SQ   Sequence   247 AA;


  Query Match           62.7%;  Score 37;  DB 20;  Length 247;
  Best Local Similarity  54.5%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      25 heakphsrpym 35


RESULT  11
AAW54091
ID   AAW54091 standard; Protein; 261 AA.
XX
AC   AAW54091;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Homo sapiens BE2 sequence.
XX
KW   BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 104
FT                   /label= Ala, Ser, Pro, Thr
FT   Misc-difference 114
FT                   /label= Gly
FT   Misc-difference 218
FT                   /label= Ala
XX
PN   WO9812327-A2.
XX
PD   26-MAR-1998.
XX
PF   19-SEP-1997;   97WO-US16842.
XX
PR   04-APR-1997;   97US-0042985.
PR   20-SEP-1996;   96US-0025296.
PR   03-APR-1997;   97US-0042611.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Baer R,  Bowcock AM;
XX
DR   WPI; 1998-230317/20.
DR   N-PSDB; AAV24126.
XX
PT   DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which
PT   as breast cancer antigen, BRCA1, binding proteins are useful to
PT   identify patient having or at risk of developing cancer
XX
PS   Disclosure; Page 272-273; 348pp; English.
XX
CC   The sequence is that of a protein which can be used in the
CC   preparation of the recombinant breast cancer antigen, BRCA1, binding
CC   proteins BARD1, B123, BE2, BE14, BE31 or BE445, or a composition for the
CC   detection of a BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid
CC   sequence, specifically a wild type BARD1 composition for the detection
CC   or purification of BRCA1, useful to identify a patient having, or at
CC   risk of developing cancer. BARD1 can be used in the preparation of an
CC   anti-BARD1 antibody, and in the detection and purification of a BRCA1
CC   protein. BARD1, B123, BE2, BE14, BE31 or BE445 can used in the
CC   identification of a binding protein agonist or antagonist that alters
CC   the binding of BARD1, B123, BE2, BE14, BE31 or BE445 to BRCA1 or the
CC   biological activity of the BRCA1-BARD1, B123, BE2, BE14, BE31 or BE445
CC   complex. The antibodies can be used to detect BARD1, B123, BE2, BE14,
CC   BE31 or BE445, a specific anti-BARD1 antibody can be used to identify
CC   a patient having or at risk of developing cancer.
XX
SQ   Sequence   261 AA;


  Query Match           62.7%;  Score 37;  DB 19;  Length 261;
  Best Local Similarity  60.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPIM 11
           ||:|||:| :
Db     219 eaqpeaqpal 228


RESULT  12
AAY90291
ID   AAY90291 standard; Protein; 267 AA.
XX
AC   AAY90291;
XX
```

DT 24-OCT-2000 (first entry)
XX
DE Human peptidase, HPEP-8 protein sequence.
XX
KW Human; peptidase; cell proliferative disorder; arteriosclerosis;
KW psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;
KW inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis;
KW Grave's disease; multiple sclerosis; scleroderma; infection; diabetes;
KW metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;
KW glycogen storage disease; obesity; therapy; HPEP-8.
XX
OS Homo sapiens.
XX
PN WO200042201-A2.
XX
PD 20-JUL-2000.
XX
PF 11-JAN-2000; 2000WO-US00641.
XX
PR 11-JAN-1999; 99US-0172247.
PR 03-MAY-1999; 99US-0132253.
PR 27-MAY-1999; 99US-0136653.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Lal P;
PI Yue H, Lu DAM;
XX
DR WPI; 2000-482832/42.
DR N-PSDB; AAA37664.
XX
PT An isolated polypeptide for diagnosis, prevention and treatment of
PT cell proliferative, autoimmune/ inflammatory and metabolic disorders
PT comprises a sequence encoding a human peptidase -
XX
PS Claim 2; Page 99; 131pp; English.
XX
CC This sequence represents a human peptidase, designated HPEP-8. The
CC invention relates to 18 human peptidases designated HPEP-1 to HPEP-18,
CC respectively. The peptidases can be used for treating a disease or
CC condition associated with decreased expression or over expression of
CC functional human peptidases. The diseases that can be diagnosed,
CC prevented and treated include cell proliferative disorders (such as
CC arteriosclerosis, psoriasis, myelofibrosis, and cancers),
CC autoimmune/inflammatory disorders (such as AIDS, anaemia, allergies,
CC Crohn's disease, asthma, atherosclerosis, Grave's disease, multiple
CC sclerosis, and scleroderma), infections, and metabolic disorders (such as
CC Addison's disease, diabetes, cystic fibrosis, glycogen storage diseases
CC and obesity).
XX
SQ Sequence 267 AA;

Query Match 62.7%; Score 37; DB 21; Length 267;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EAEPEAEP 9
| ||||||
Db 247 epepeaep 254

RESULT 13
AAB20156
ID AAB20156 standard; Protein; 267 AA.
XX
AC AAB20156;
XX
DT 30-APR-2001 (first entry)
XX
DE Human protein SECP2.
XX
KW SECP2; secreted protein; human; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN WO200105971-A2.
XX
PD 25-JAN-2001.
XX
PF 20-JUL-2000; 2000WO-US19890.
XX
PR 20-JUL-1999; 99US-0144722.
PR 29-NOV-1999; 99US-0167785.
PR 19-JUL-2000; 2000US-0619252.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Fernandes E;
XX
DR WPI; 2001-091973/10.
DR N-PSDB; AAF30189.
XX
PT New polypeptide designated SECP, its encoding nucleic acid and its
PT immunospecific antibody, useful for diagnosing, preventing and treating
PT SECP-associated disorders such as cancer -
XX
PS Claim 1; Fig 2; 124pp; English.
XX
CC The present sequence is that of novel human protein SECP2,
CC which is predicted to localise in the microbody (peroxisome),
CC and which does not appear to include a signal peptide. The
CC protein shows homology to human PRO351 protein, and to a region
CC of human prostatin precursor. The invention provides 9 novel
CC SECP proteins (see AAB20155-63), nucleic acids encoding them
CC (see AAF30188-96), antibodies, mutants or fragments. These can
CC be used to detect, treat or prevent an SECP-associated disorder,
CC to screen for predisposition to such a disorder, and to identify
CC agents that modulate the expression or activity of SECP.
XX
SQ Sequence 267 AA;

Query Match 62.7%; Score 37; DB 22; Length 267;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EAEPEAEP 9
| ||||||
Db 247 epepeaep 254

RESULT 14
AAB58142
ID AAB58142 standard; Protein; 278 AA.
XX
AC AAB58142;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polypeptide sequence SEQ ID 480.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX

```
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18018.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 961-962; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   278 AA;
```

```
  Query Match             62.7%;  Score 37;  DB 21;  Length 278;
  Best Local Similarity   54.5%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 HEAEPEAEPIM 11
           |||:| : | |
Db      57 heakphsrpym 67
```

```
RESULT  15
AAE00235
ID   AAE00235 standard; Protein; 302 AA.
XX
AC   AAE00235;
XX
DT   13-JUN-2001  (first entry)
XX
DE   Protein encoded by Probe 9.
XX
KW   Transacylase; taxol; paclitaxel biosynthesis; taxoid; probe;
KW   transgenic organism; TAX9.
XX
OS   Taxus cuspidata.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 119
FT                   /note= "Encoded by TGG"
FT   Misc-difference 164
FT                   /note= "Encoded by ATS"
FT   Misc-difference 187
FT                   /note= "Encoded by TGG"
XX
PN   WO200123586-A2.
XX
PD   05-APR-2001.
XX
PF   29-SEP-2000; 2000WO-US27006.
XX
PR   30-SEP-1999;   99US-0411145.
PR   07-DEC-1999;   99US-0457046.
XX
PA   (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
PI   Croteau RB,  Walker KD,  Schoendorf A,  Wildung MR;
XX
DR   WPI; 2001-245004/25.
DR   N-PSDB; AAD03341.
XX
PT   New transacylase enzymes, useful for the high yield production of
PT   Taxol(TM), related taxoids and useful intermediates in the in the
PT   paclitaxel biosynthetic pathway  -
XX
PS   Claim 1; Page 108-109; 162pp; English.
XX
CC   The present sequence is the amino acid sequence of probe 9.
CC   Probe 9 is derived from AT-FOR2 and AT-REV1 primers and is
CC   used for screening Taxus cuspidata TAX9 full length cDNA clone. The
CC   probes are useful for the identification of (nucleic acid sequences
CC   encoding) transacylases. The probes isolated from the Taxus genus
CC   are useful for the synthetic production of Taxol(TM) and related taxoids,
CC   as well as intermediates in the paclitaxel biosynthetic pathway. They
CC   can also be used for the creation of transgenic organisms that either
CC   produce the transacylases for subsequent in vitro use, or produce the
CC   transacylases in vivo. The (nucleic acids encoding) transacylases are
CC   also useful for creating specific binding agents that recognise the
CC   corresponding transacylases. Binding agents include (fragments of)
CC   antibodies or any other agent capable of specifically binding to the
CC   groups on the proteins.
XX
SQ   Sequence   302 AA;
```

```
  Query Match             62.7%;  Score 37;  DB 22;  Length 302;
  Best Local Similarity   70.0%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 EAEPEAEPIM 11
           ||:|  ||||
Db     110 eakpslepim 119
```

```
Search completed: January  4, 2002, 08:40:26
Job time: 109 sec
```

```
OM of: US-09-444-281-35 to: EST:*   out_format : pfs

Date: Jan 4, 2002 10:25 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084143_16183/app_query.fasta_1.210
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09444281_@CGN1_1_3354 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-35
Query length: 13
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 2868.270000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_est1:BE024584        +    69.00  146.30   55.12     330  ! BE024584 894003H02.y1 C. reinha
gb_gss:AZ518927         -    69.00  145.09   64.43     390  ! AZ518927 RPCI-11-67B18.TV RPCI-
gb_est1:BE237369        +    68.00  142.66   87.89     415  ! BE237369 146629 MARC 4BOV Bos t
gb_est1:AU089922        -    68.00  142.14   94.01     446  ! AU089922 AU089922 Hordeum vulga
gb_est1:AU198144        -    68.00  142.14   94.01     446  ! AU198144 AU198144 Rice green sh
gb_est1:AU198162        -    68.00  142.11   94.40     448  ! AU198162 AU198162 Rice green sh
gb_est1:AU089934        -    68.00  141.68   99.71     475  ! AU089934 AU089934 Hordeum vulga
gb_est1:AU082117        -    68.00  140.25  119.78     578  ! AU082117 AU082117 Rice panicle
gb_est1:AU198258        -    66.00  139.93  124.83     352  ! AU198258 AU198258 Rice green sh
gb_est1:BE453064        -    66.00  138.40  151.82     434  ! BE453064 894068B05.x1 C. reinha
gb_est2:BG305131        -    66.00  138.22  155.41     445  ! BG305131 fl95a09.y1 Zebrafish a
gb_gss:AZ368740         -    66.00  136.04  205.48     600  ! AZ368740 1M0118O21R Mouse 10kb
gb_est2:BE977776        +    65.00  140.31  118.88     255  ! BE977776 bs68h10.y1 Drosophila
gb_est1:AV641634        +    65.00  137.58  168.76     371  ! AV641634 AV641634 Chlamydomonas
gb_est1:BE129188        +    65.00  137.29  175.13     386  ! BE129188 894021E12.y1 C. reinha
gb_est1:AU090148        +    65.00  136.34  197.93     440  ! AU090148 AU090148 Hordeum vulga
gb_est1:AU090136        +    65.00  136.04  205.48     458  ! AU090136 AU090136 Hordeum vulga
gb_est2:BE978206        +    65.00  134.78  241.74     545  ! BE978206 bs75d09.y1 Drosophila
gb_est1:AW762881        -    64.00  133.42  287.61     501  ! AW762881 ur59b09.y1 NCI_CGAP_Ma
gb_est2:C52835          +    63.00  137.86  162.80     208  ! C52835 C52835 Yuji Kohara unpub
gb_est2:BG051614        -    63.00  135.61  217.08     283  ! BG051614 FM1_58_H09.g1_A003 Flo
gb_est2:C62320          -    63.00  135.39  223.52     292  ! C62320 C62320 Yuji Kohara unpub
gb_est1:BE494459        +    63.00  132.79  311.84     417  ! BE494459 WHE1254_A02_B04ZS Seca
gb_gss:AQ448482         -    63.00  131.11  386.73     525  ! AQ448482 mgxb0020D20f CUGI Rice
gb_est2:BI159202        -    63.00  130.23  433.34     593  ! BI159202 602922713F1 NIH_CGAP_M
gb_est2:BI174860        +    63.00  130.14  438.12     600  ! BI174860 OSTF061C5_1 AD-wrmcDNA
gb_est2:BF584442        +    63.00  129.01  506.68     701  ! BF584442 602098336F1 NCI_CGAP_C
gb_gss:CNS02PFP         -    63.00  128.86  516.13     715  ! AL207934 Tetraodon nigroviridis
gb_est2:BG964576        +    63.00  127.74  595.99     834  ! BG964576 602832255F1 NCI_CGAP_C
gb_gss:CNS04848         +    63.00  127.64  603.34     845  ! AL278801 Tetraodon nigroviridis
gb_est2:BG711576        +    62.00  130.11  439.92     460  ! BG711576 pg11n.pk008.f16 Normal
gb_est2:BG713047        +    62.00  129.56  472.01     496  ! BG713047 pg11n.pk015.c1 Normali
gb_est1:AW198314        +    62.00  129.51  474.68     499  ! AW198314 pnl-b.pk0005.d1 chicke
gb_est2:BG712658        +    62.00  128.73  525.17     556  ! BG712658 pg11n.pk013.g21 Normal
gb_est2:BG711313        +    62.00  127.92  582.33     621  ! BG711313 pg11n.pk007.c13 Normal
gb_est1:BE123173        -    61.50  132.56  321.16     287  ! BE123173 945024C07.Y2 945 - Mix
gb_est1:BE123317        -    61.50  129.37  483.86     445  ! BE123317 945041B02.Y1 945 - Mix
gb_est1:AW424548        -    61.50  129.24  491.99     453  ! AW424548 707017F11.y1 707 - Mix
gb_est2:BG901201        -    61.50  129.16  497.06     458  ! BG901201 950001B11.x2 950 - Mat
gb_est1:BE643645        -    61.50  128.83  518.33     479  ! BE643645 945041B06.Y2 945 - Mix
gb_est1:AW562908        -    61.50  128.47  542.56     503  ! AW562908 660069E08.x1 660 - Mix
gb_est1:BE515909        -    61.50  128.40  547.60     508  ! BE515909 WHE0604_D03_H06ZA W
gb_est1:BE123319        -    61.50  128.30  554.64     515  ! BE123319 945041B06.Y1 945 -
gb_est1:AW562591        +    61.50  128.19  562.69     523  ! AW562591 660066G07.y1 660 -
gb_est1:BE643603        -    61.50  128.12  567.72     528  ! BE643603 945024C07.Y3 945 -

seq_name: gb_est1:BE024584

seq_documentation_block:
LOCUS       BE024584      330 bp    mRNA            EST       06-JUN-2000
DEFINITION  894003H02.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE024584
VERSION     BE024584.1  GI:8287025
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT       40 a    103 c    142 g     45 t
ORIGIN


alignment_scores:
            Quality:   69.00              Length:     11
              Ratio:  6.900                Gaps:      0
 Percent Similarity:  90.909   Percent Identity: 72.727

alignment_block:
 US-09-444-281-35 x BE024584   ..

 Align seg 1/1  to: BE024584  from: 1  to: 330

       2 LeuLysLysTrpProTrpTrpProTrpArgArg 12
         |||:::::|||||||||||||||||||   |||
      66 CTAGAACGGTGGCCGTGGTGGCCGTGGGCGCGG 98

seq_name: gb_gss:AZ518927

seq_documentation_block:
```

```
LOCUS       AZ518927      390 bp    DNA             GSS       16-OCT-2000
DEFINITION  RPCI-11-67B18.TV RPCI-11 Homo sapiens genomic clone RPCI-11-67B18,
            DNA sequence.
ACCESSION   AZ518927
VERSION     AZ518927.1  GI:10829921
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     BAC end sequences of library RPCI-11
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-67B18.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            This BAC end was generated during the R&D process and may have
            higher chance of clone tracking errors.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Homo sapiens"
                     /db_xref="GDB:7525385"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-67B18"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       59 a    159 c     84 g     87 t      1 others
ORIGIN


alignment_scores:
            Quality:   69.00            Length:     12
              Ratio:  6.273              Gaps:      0
 Percent Similarity:  91.667   Percent Identity: 83.333

alignment_block:
 US-09-444-281-35 x AZ518927/rev   ..

 Align seg 1/1  to reverse of: AZ518927  from: 1  to: 390

       2 LeuLysLysTrpProTrpTrpProTrpArgArgLys 13
         |||:::|||   ||||||||||||||||||||||||
     189 TTACAAAAATCCCCCTGGTGGCCCTGGAGGAGGAAG 154

seq_name: gb_est1:BE237369

seq_documentation_block:
LOCUS       BE237369      415 bp    mRNA            EST       25-APR-2001
DEFINITION  146629 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE237369
VERSION     BE237369.1  GI:9022087
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 47  row: J  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT      106 a     97 c    120 g     91 t      1 others
ORIGIN


alignment_scores:
            Quality:   68.00            Length:      8
              Ratio:  8.500              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x BE237369   ..

 Align seg 1/1  to: BE237369  from: 1  to: 415

       4 LysTrpProTrpTrpProTrpArg 11
         |||||||||||||||||||||||||
     280 AAATGGCCATGGTGGCCTTGGCGC 303

seq_name: gb_est1:AU089922

seq_documentation_block:
LOCUS       AU089922      446 bp    mRNA            EST       19-APR-2000
DEFINITION  AU089922 Hordeum vulgare subsp. vulgare Upper three leaves at
            heading stage Hordeum vulgare subsp. vulgare cDNA clone
            haruna_lib1_121, mRNA sequence.
ACCESSION   AU089922
VERSION     AU089922.1  GI:7613350
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Sato,K., Takahashi,H. and Takeda,K.
  TITLE     Hordeum vulgare subsp. vulgare cDNA clone
  JOURNAL   Unpublished (2000)
```

```
COMMENT     Contact: Kazuhiro Sato
            Research Institute for Bioresources
            Okayama University, Barley Germplasm Center
            Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
            Email: kazsato@rib.okayama-u.ac.jp,
            URL:http://www.rib.okayama-u.ac.jp/barley/.
FEATURES              Location/Qualifiers
     source           1. .446
                      /organism="Hordeum vulgare subsp. vulgare"
                      /cultivar="Haruna Nijo"
                      /db_xref="taxon:112509"
                      /clone="haruna_lib1_121"
                      /clone_lib="Hordeum vulgare subsp. vulgare Upper three
                      leaves at heading stage"
                      /tissue_type="Upper three leaves at heading stage"
BASE COUNT       89 a    130 c    149 g     76 t      2 others
ORIGIN


alignment_scores:
           Quality:   68.00             Length:      8
             Ratio:  8.500               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AU089922/rev   ..

 Align seg 1/1  to reverse of: AU089922  from: 1  to: 446

       5 TrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||
     188 TGGCCGTGGTGGCCGTGGCGGCGA 165

seq_name: gb_est1:AU198144

seq_documentation_block:
LOCUS       AU198144      446 bp    mRNA            EST       12-JUL-2001
DEFINITION  AU198144 Rice green shoot Oryza sativa cDNA clone S15951, mRNA
            sequence.
ACCESSION   AU198144
VERSION     AU198144.1  GI:14714211
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT = 'RGP'.
            S15951_97A.
FEATURES              Location/Qualifiers
     source           1. .446
                      /organism="Oryza sativa"
                      /strain="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone="S15951"
                      /clone_lib="Rice green shoot"
                      /note="Green shoot (8 days old)"
BASE COUNT       82 a    153 c    161 g     50 t
ORIGIN


alignment_scores:
           Quality:   68.00             Length:      11
             Ratio:  6.800               Gaps:       0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-35 x AU198144/rev   ..

 Align seg 1/1  to reverse of: AU198144  from: 1  to: 446

       3 LysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::||||||||||||||||||   |||:::
     351 CGCCGCTGGCCTTGGTGGCCCTGGACGCGGCGG 319

seq_name: gb_est1:AU198162

seq_documentation_block:
LOCUS       AU198162      448 bp    mRNA            EST       12-JUL-2001
DEFINITION  AU198162 Rice green shoot Oryza sativa cDNA clone S16019, mRNA
            sequence.
ACCESSION   AU198162
VERSION     AU198162.1  GI:14714231
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT = 'RGP'.
            S16019_97A.
FEATURES              Location/Qualifiers
     source           1. .448
                      /organism="Oryza sativa"
                      /strain="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone="S16019"
                      /clone_lib="Rice green shoot"
                      /note="Green shoot (8 days old)"
BASE COUNT       85 a    146 c    160 g     57 t
ORIGIN


alignment_scores:
           Quality:   68.00             Length:      11
             Ratio:  6.800               Gaps:       0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-35 x AU198162/rev   ..

 Align seg 1/1  to reverse of: AU198162  from: 1  to: 448

       3 LysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::||||||||||||||||||   |||:::
     314 CGCCGCTGGCCTTGGTGGCCCTGGACGCGGCGG 282

seq_name: gb_est1:AU089934

seq_documentation_block:
LOCUS       AU089934      475 bp    mRNA            EST       19-APR-2000
DEFINITION  AU089934 Hordeum vulgare subsp. vulgare Upper three leaves at
            heading stage Hordeum vulgare subsp. vulgare cDNA clone
            haruna_lib1_134, mRNA sequence.
```

```
ACCESSION   AU089934
VERSION     AU089934.1  GI:7613362
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Sato,K., Takahashi,H. and Takeda,K.
  TITLE     Hordeum vulgare subsp. vulgare cDNA clone
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Kazuhiro Sato
            Research Institute for Bioresources
            Okayama University, Barley Germplasm Center
            Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
            Email: kazsato@rib.okayama-u.ac.jp,
            URL:http://www.rib.okayama-u.ac.jp/barley/.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="haruna_lib1_134"
                     /clone_lib="Hordeum vulgare subsp. vulgare Upper three
                     leaves at heading stage"
                     /tissue_type="Upper three leaves at heading stage"
BASE COUNT      100 a    137 c    155 g     83 t
ORIGIN


alignment_scores:
            Quality:   68.00             Length:      8
              Ratio:  8.500               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-35 x AU089934/rev   ..

 Align seg 1/1  to reverse of: AU089934  from: 1  to: 475

       5 TrpProTrpTrpProTrpArgArg 12
         ||||||||||||||||||||||||
     194 TGGCCGTGGTGGCCGTGGCGGCGA 171

seq_name: gb_est1:AU082117

seq_documentation_block:
LOCUS       AU082117      578 bp    mRNA            EST       04-FEB-2000
DEFINITION  AU082117 Rice panicle at ripening stage Oryza sativa cDNA clone
            E11611, mRNA sequence.
ACCESSION   AU082117
VERSION     AU082117.1  GI:6727452
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle at ripening stage (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E11611"
                     /clone_lib="Rice panicle at ripening stage"
                     /dev_stage="ripening stage"
                     /note="Organ: panicle; Rice cDNA from panicle at ripening
                     stage"
BASE COUNT      129 a    180 c    179 g     90 t
ORIGIN


alignment_scores:
            Quality:   68.00             Length:      11
              Ratio:  6.800               Gaps:       0
 Percent Similarity:  90.909   Percent Identity: 63.636

alignment_block:
 US-09-444-281-35 x AU082117/rev   ..

 Align seg 1/1  to reverse of: AU082117  from: 1  to: 578

       3 LysLysTrpProTrpTrpProTrpArgArgLys 13
         ::::::||||||||||||||||||   |||:::
     348 CGCCGCTGGCCTTGGTGGCCCTGGACGCGGCGG 316

seq_name: gb_est1:AU198258

seq_documentation_block:
LOCUS       AU198258      352 bp    mRNA            EST       12-JUL-2001
DEFINITION  AU198258 Rice green shoot Oryza sativa cDNA clone S16389, mRNA
            sequence.
ACCESSION   AU198258
VERSION     AU198258.1  GI:14714335
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT = 'RGP'.
            S16389_96Z.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S16389"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT       95 a     95 c    104 g     51 t      7 others
ORIGIN


alignment_scores:
            Quality:   66.00             Length:      10
              Ratio:  7.333               Gaps:       0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-35 x AU198258/rev   ..
```

```
 Align seg 1/1  to reverse of: AU198258  from: 1  to: 352

       3 LysLysTrpProTrpTrpProTrpArgArg 12
         ::::::||||||||||||||||||||   |||
     170 CGCCGCTGGCCTTGGTGGCCCTGGACGCGG 141

seq_name: gb_est1:BE453064

seq_documentation_block:
LOCUS       BE453064      434 bp    mRNA            EST       26-JUL-2000
DEFINITION  894068B05.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE453064
VERSION     BE453064.1  GI:9460101
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      110 a    146 c     86 g     92 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     11
              Ratio:  7.333               Gaps:      0
 Percent Similarity:  81.818   Percent Identity: 72.727

alignment_block:
 US-09-444-281-35 x BE453064/rev   ..

 Align seg 1/1  to reverse of: BE453064  from: 1  to: 434

       1 IleLeuLysLysTrpProTrpTrpProTrpArg 11
         |||:::     ||||||||||||||||||||||
     403 ATAATGGTGGCATGGCCATGGTGGCCATGGCGT 371

seq_name: gb_est2:BG305131

seq_documentation_block:
LOCUS       BG305131      445 bp    mRNA            EST       22-FEB-2001
DEFINITION  fl95a09.y1 Zebrafish adult retina cDNA Danio rerio cDNA clone
            4144409 5', mRNA sequence.
ACCESSION   BG305131
VERSION     BG305131.1  GI:13102658
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone="4144409"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT       97 a    140 c    119 g     89 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     11
              Ratio:  7.333               Gaps:      0
 Percent Similarity:  81.818   Percent Identity: 72.727

alignment_block:
 US-09-444-281-35 x BG305131/rev   ..

 Align seg 1/1  to reverse of: BG305131  from: 1  to: 445

       3 LysLysTrpProTrpTrpProTrpArgArgLys 13
         |||   ||||||||||||||||||   |||:::
     252 AAACTGTGGCCCTGGTGGCCATGGAGCAGAGAG 220

seq_name: gb_gss:AZ368740

seq_documentation_block:
LOCUS       AZ368740      600 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0118O21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0118O21 R, DNA sequence.
ACCESSION   AZ368740
VERSION     AZ368740.1  GI:10482440
```

```
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0118  row: O  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0118O21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      173 a    108 c    181 g    138 t
ORIGIN


alignment_scores:
            Quality:   60.00             Length:      9
              Ratio:  7.333              Gaps:        0
 Percent Similarity: 100.000  Percent Identity: 77.778

alignment_block:
 US-09-444-281-35 x AZ368740/rev  ..

 Align seg 1/1  to reverse of: AZ368740  from: 1  to: 600

       2 LeuLysLysTrpProTrpTrpProTrp 10
         |||::::::|||||||||||||||||||
     384 TTACGGAGGTGGCCATGGTGGCCCTGG 358

seq_name: gb_est2:BE977776
```

```
seq_documentation_block:
LOCUS       BE977776      255 bp    mRNA            EST       04-OCT-2000
DEFINITION  bs68h10.y1 Drosophila melanogaster adult testis library Drosophila
            melanogaster cDNA clone bs68h10 5', mRNA sequence.
ACCESSION   BE977776
VERSION     BE977776.1  GI:10608588
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 255)
  AUTHORS   Andrews, J., Bouffard, G. and Oliver, B.
  TITLE     Drosophila melanogaster testis expressed sequence tags
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Brian Oliver
            Laboratory of Cellular and Developmental Biology
            NIDDK, National Institutes of Health
            6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
            Fax: (301) 496 5239
            Email: oliver@helix.nih.gov,
            http://www.niddk.nih.gov/intram/people/boliver.htm
            Tissue isolation and library construction performed at the National
            Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
            http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
            and analyses performed by National Institutes of Health Intramural
            Sequencing Center (NISC; see http://www.nisc.nih.gov).
            Plate: 68  row: h  column: 10
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .255
                     /organism="Drosophila melanogaster"
                     /strain="y[*] w[67c1]/Y"
                     /db_xref="taxon:7227"
                     /clone="bs68h10"
                     /clone_lib="Drosophila melanogaster adult testis library"
                     /sex="male"
                     /dev_stage="1-5 day adult"
                     /lab_host="SOLR (Stratagene)"
                     /note="Organ: testis; Vector: pBlueScript SK (Stratagene);
                     Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
                     day adult y[*] w[67c1]/Y males raised at 25oC.  RNA
                     isolated using Trizol (Life Technologies) and a single
                     round of Poly(A)+ selection using Oligotex (Qiagen).  cDNA
                     library constructed using Stratagene ZAP-cDNA syntesis
                     kit.  Oligo dT-primed, size fractionated ~1-6 kb, and
                     directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
                     Following a single round of amplification pBlueScript SK
                     phagemids were mass excised.  A distribution channel for
                     clones is being sought, but not currently available.
                     Requests for clones cannot be honored."
BASE COUNT       47 a     60 c    101 g     47 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     15
              Ratio:  5.909              Gaps:        1
 Percent Similarity:  73.333  Percent Identity: 60.000

alignment_block:
 US-09-444-281-35 x BE977776  ..

 Align seg 1/1  to: BE977776  from: 1  to: 255

       3 LysLysTrpPro...........TrpTrpProTrpArgArgLys 13
         |||:::||||||           ||||||||||||||:::|||
     185 AAGAGGTGGCCCTGGAGGAAAAGGTGGTGGCCCTGGAGGAAGAAA 229

seq_name: gb_est1:AV641634
```

```
seq_documentation_block:
LOCUS       AV641634      371 bp    mRNA            EST       15-DEC-2000
DEFINITION  AV641634 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HCL037h03_r 5', mRNA sequence.
ACCESSION   AV641634
VERSION     AV641634.1  GI:10784962
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HCL037h03_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       73 a    106 c    141 g     51 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     10
              Ratio:  7.222               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-35 x AV641634   ..

 Align seg 1/1  to: AV641634  from: 1  to: 371

       3 LysLysTrpProTrpTrpProTrpArgArg 12
         ::::::||||||||||||||||||   |||
     241 GAACGGTGGCCGTGGTGGCCGTGGGCGCGG 270

seq_name: gb_est1:BE129188

seq_documentation_block:
LOCUS       BE129188      386 bp    mRNA            EST       21-JUN-2000
DEFINITION  894021E12.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE129188
VERSION     BE129188.1  GI:8576551
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT       79 a    109 c    142 g     56 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     10
              Ratio:  7.222               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-35 x BE129188   ..

 Align seg 1/1  to: BE129188  from: 1  to: 386

       3 LysLysTrpProTrpTrpProTrpArgArg 12
         ::::::||||||||||||||||||   |||
     246 GAACGGTGGCCGTGGTGGCCGTGGGCGCGG 275
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:47:23 ; Search time 50.17 Seconds
(without alignments)
37.902 Million cell updates/sec

Title: US-09-444-281-35
Perfect score: 91
Sequence: 1 ILKKWPWWPWRRK 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	57	62.6	723	12	Q9DUC4	Q9duc4 tt virus. o
2	54	59.3	1173	12	Q990M4	Q990m4 human coron
3	54	59.3	1173	12	Q990M3	Q990m3 human coron
4	54	59.3	1173	12	Q990M2	Q990m2 human coron
5	54	59.3	1173	12	Q990M1	Q990m1 human coron
6	54	59.3	1383	12	Q84712	Q84712 porcine epi
7	52	57.1	746	12	Q9JH31	Q9jh31 tt virus. o
8	52	57.1	1018	1	Q9HKX3	Q9hkx3 thermoplasm
9	51	56.0	299	4	Q9Y4N1	Q9y4n1 homo sapien
10	50	54.9	327	10	Q9AUN3	Q9aun3 oryza sativ
11	49	53.8	148	5	Q26590	Q26590 schistosoma
12	49	53.8	298	1	Q9Y8Q6	Q9y8q6 aeropyrum p
13	49	53.8	467	5	Q19573	Q19573 caenorhabdi
14	49	53.8	528	5	Q26589	Q26589 schistosoma
15	49	53.8	528	5	Q9TY57	Q9ty57 schistosoma
16	49	53.8	735	12	Q9DUC9	Q9duc9 tt virus. o
17	49	53.8	802	5	O96398	O96398 schistosoma
18	49	53.8	1245	3	Q9Y7V5	Q9y7v5 trichoderma
19	49	53.8	1940	5	Q02456	Q02456 schistosoma
20	48	52.7	49	12	Q9DT80	Q9dt80 tt virus. o
21	48	52.7	111	5	Q18753	Q18753 caenorhabdi
22	48	52.7	428	11	Q9JMG0	Q9jmg0 mus musculu
23	48	52.7	431	11	Q99ML4	Q99ml4 mus musculu
24	48	52.7	748	12	Q9DT81	Q9dt81 tt virus. o
25	47.5	52.2	114	2	Q9X8C2	Q9x8c2 streptomyce
26	47	51.6	141	11	Q9CZA1	Q9cza1 mus musculu
27	47	51.6	165	10	Q9SNN3	Q9snn3 oryza sativ
28	47	51.6	504	2	P96143	P96143 thermoactin
29	46.5	51.1	352	2	P73417	P73417 synechocyst
30	46.5	51.1	620	12	Q9IH07	Q9ih07 avian infec
31	46.5	51.1	621	12	Q66196	Q66196 avian infec
32	46.5	51.1	621	12	Q9IH14	Q9ih14 avian infec
33	46.5	51.1	621	12	Q9IH13	Q9ih13 avian infec
34	46.5	51.1	621	12	Q9IH11	Q9ih11 avian infec
35	46.5	51.1	621	12	Q9IH10	Q9ih10 avian infec
36	46.5	51.1	621	12	Q9IH15	Q9ih15 avian infec
37	46.5	51.1	621	12	Q9IH12	Q9ih12 avian infec
38	46.5	51.1	621	12	Q9IH09	Q9ih09 avian infec
39	46.5	51.1	621	12	Q9IH08	Q9ih08 avian infec
40	46.5	51.1	621	12	Q991L8	Q99118 avian infec
41	46.5	51.1	621	12	Q991L7	Q99117 avian infec
42	46.5	51.1	621	12	Q991L6	Q99116 avian infec
43	46.5	51.1	625	12	Q9QCP6	Q9qcp6 avian infec
44	46.5	51.1	630	12	Q66197	Q66197 avian infec
45	46.5	51.1	630	12	Q98WQ0	Q98wq0 avian infec

ALIGNMENTS

RESULT 1
Q9DUC4
ID Q9DUC4 PRELIMINARY; PRT; 723 AA.
AC Q9DUC4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RA Okamoto H.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RX PubMed=11080484;
RA Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,
RA Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;
RT "Species-specific TT viruses in humans and nonhuman primates and their
RT phylogenetic relatedness.";
RL Virology 277:368-378(2000).
DR EMBL; AB041959; BAB19313.1; -.
DR InterPro; IPR001563; Serine_carbpept.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; UNKNOWN_1.
SQ SEQUENCE 723 AA; 85393 MW; 232D003098766344 CRC64;

Query Match 62.6%; Score 57; DB 12; Length 723;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PWWPWRR 12
|||||||
Db 2 PWWPWRR 8

RESULT 2
Q990M4

```
ID   Q990M4      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344186; AAK32188.1; -.
SQ   SEQUENCE   1173 AA;  128669 MW;  ABC6E0A75E8BD8A4 CRC64;


  Query Match             59.3%;  Score 54;  DB 12;  Length 1173;
  Best Local Similarity   85.7%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 KWPWWPW 10
          ||||| |
Db   1113 KWPWWVW 1119


RESULT   3
Q990M3
ID   Q990M3      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344187; AAK32189.1; -.
SQ   SEQUENCE   1173 AA;  128683 MW;  9E2368160082A81A CRC64;


  Query Match             59.3%;  Score 54;  DB 12;  Length 1173;
  Best Local Similarity   85.7%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 KWPWWPW 10
          ||||| |
Db   1113 KWPWWVW 1119


RESULT   4
Q990M2
ID   Q990M2      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344188; AAK32190.1; -.
SQ   SEQUENCE   1173 AA;  128653 MW;  8B658FCBBD1842DA CRC64;


  Query Match             59.3%;  Score 54;  DB 12;  Length 1173;
  Best Local Similarity   85.7%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 KWPWWPW 10
          ||||| |
Db   1113 KWPWWVW 1119


RESULT   5
Q990M1
ID   Q990M1      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344189; AAK32191.1; -.
SQ   SEQUENCE   1173 AA;  128760 MW;  B73A165A6270152A CRC64;


  Query Match             59.3%;  Score 54;  DB 12;  Length 1173;
  Best Local Similarity   85.7%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 KWPWWPW 10
          ||||| |
Db   1113 KWPWWVW 1119


RESULT   6
Q84712
ID   Q84712      PRELIMINARY;      PRT;  1383 AA.
AC   Q84712;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE PROTEIN.
OS   porcine epidemic diarrhea virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=28295;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1/87;
RX   MEDLINE=94231173; PubMed=8176382;
```

```
RA   Duarte M., Laude H.;
RT   "Sequence of the spike protein of the porcine epidemic diarrhoea
RT   virus.";
RL   J. Gen. Virol. 75:1195-1200(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1/87;
RX   MEDLINE=93389433; PubMed=8397280;
RA   Bridgen A., Duarte M., Tobler K., Laude H., Ackermann M.;
RT   "Sequence determination of the nucleocapsid protein gene of the
RT   porcine epidemic diarrhoea virus confirms that this virus is a
RT   coronavirus related to human coronavirus 229E and porcine
RT   transmissible gastroenteritis virus.";
RL   J. Gen. Virol. 74:1795-1804(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1/87;
RX   MEDLINE=94120721; PubMed=8291230;
RA   Duarte M., Tobler K., Bridgen A., Rasschaert D., Ackermann M.,
RA   Laude H.;
RT   "Sequence analysis of the porcine epidemic diarrhea virus genome
RT   between the nucleocapsid and spike protein genes reveals a polymorphic
RT   ORF.";
RL   Virology 198:466-476(1994).
DR   EMBL; Z25483; CAA80971.1; -.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
FT   CONFLICT    422    422       Y -> N (IN REF. 1).
SQ   SEQUENCE   1383 AA;  151404 MW;  741C84D5DD3BDC4D CRC64;


  Query Match             59.3%;  Score 54;  DB 12;  Length 1383;
  Best Local Similarity   85.7%;  Pred. No. 18;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 KWPWWPW 10
          ||||| |
Db   1322 KWPWWVW 1328


RESULT   7
Q9JH31
ID   Q9JH31      PRELIMINARY;      PRT;   746 AA.
AC   Q9JH31;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ORF1.
OS   TT virus.
OC   Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX   NCBI_TaxID=68887;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TJN02;
RA   Okamoto H.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TJN02;
RA   Ukita M., Okamoto H., Nishizawa T., Tawara A., Takahashi M.,
RA   Iizuka H., Miyakawa Y., Mayumi M.;
RT   "The entire nucleotide sequences of two distinct TT virus (TTV)
RT   isolates (TJN01 and TJN02) remotely related to the original TTV
RT   isolates.";
RL   Arch. Virol. 0:0-0(2000).
DR   EMBL; AB028669; BAA94878.1; -.
SQ   SEQUENCE   746 AA;  88561 MW;  E0B22953AE764E3E CRC64;


  Query Match             57.1%;  Score 52;  DB 12;  Length 746;
  Best Local Similarity   66.7%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 WPWWPWRRK 13
          | || |||:
Db      3 WGWWRWRRR 11


RESULT   8
Q9HKX3
ID   Q9HKX3      PRELIMINARY;      PRT;  1018 AA.
AC   Q9HKX3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CONSERVED HYPOTHETICAL MEMBRANE PROTEIN.
GN   TA0470.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445064; CAC11612.1; -.
DR   InterPro; IPR000731; HMGCR_patched_5TM.
DR   PROSITE; PS50156; SSD; 1.
KW   Complete proteome.
SQ   SEQUENCE   1018 AA;  112323 MW;  83EE84D33C74B852 CRC64;


  Query Match             57.1%;  Score 52;  DB 1;  Length 1018;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWWP 9
          ::||| |||
Db   1004 LMKKWNWWP 1012


RESULT   9
Q9Y4N1
ID   Q9Y4N1      PRELIMINARY;   PRT;   299 AA.
AC   Q9Y4N1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 34.0 KDA PROTEIN (FRAGMENT).
GN   DKFZP434C192.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Ansorge W., Wirkner U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL096753; CAB46428.2; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   299 AA;  34032 MW;  6B8DB60E6A88239A CRC64;


  Query Match             56.0%;  Score 51;  DB 4;  Length 299;
  Best Local Similarity   85.7%;  Pred. No. 11;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      6 PWWPWRR 12
          ||||| |
Db     37 PWWPWSR 43


RESULT  10
Q9AUN3
ID   Q9AUN3      PRELIMINARY;      PRT;   327 AA.
AC   Q9AUN3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Spiegel L.A., King L., Kirchoff K.A., de la Bastide M., Preston R.R.,
RA   Nascimento L.U., Vil M.D., Baker J.P., Miller B., Cunnius D.M.,
RA   Kuit K.H., Rodriguez S., Santos L., Zutavern T., Balija V.S.,
RA   Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
RA   McCombie W.R.;
RT   "Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
RT   Clone OSJNBa0058E19, Complete Sequence.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC083945; AAK13143.1; -.
SQ   SEQUENCE   327 AA;  36672 MW;  5CCA9080664BD0CA CRC64;


  Query Match             54.9%;  Score 50;  DB 10;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 WWPWRR 12
          ||||||
Db    119 WWPWRR 124


RESULT  11
Q26590
ID   Q26590      PRELIMINARY;      PRT;   148 AA.
AC   Q26590;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   MYOSIN HEAVY CHAIN (FRAGMENT).
OS   Schistosoma mansoni (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC   Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae;
OC   Schistosoma.
OX   NCBI_TaxID=6183;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schmitz J., Symmons P., Dargatz H., Kunz W.;
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M81338; AAA29906.1; -.
KW   Myosin.
FT   NON_TER       1      1
FT   NON_TER     148    148
SQ   SEQUENCE   148 AA;  17923 MW;  C7EDA5A0BBE14DDA CRC64;


  Query Match             53.8%;  Score 49;  DB 5;  Length 148;
  Best Local Similarity   62.5%;  Pred. No. 11;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWW 8
          :|: ||||
Db     30 VLRNWPWW 37


RESULT  12
Q9Y8Q6
ID   Q9Y8Q6      PRELIMINARY;      PRT;   298 AA.
AC   Q9Y8Q6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 33.7 KDA PROTEIN APE2577.
GN   APE2577.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000064; BAA81594.1; -.
DR   InterPro; IPR002787; DUF85.
DR   Pfam; PF01932; DUF85; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   298 AA;  33666 MW;  FCBE9C6EC93FF231 CRC64;


  Query Match             53.8%;  Score 49;  DB 1;  Length 298;
  Best Local Similarity   60.0%;  Pred. No. 21;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 LKKWPWWPWR 11
          :|: ||| ||
Db    102 IKETPWWRWR 111


RESULT  13
Q19573
ID   Q19573      PRELIMINARY;      PRT;   467 AA.
AC   Q19573;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   SIMILARITY TO 9 AMINO ACID REPEATS IN GALACTOSE SPECIFIC LECTINS.
GN   F18G5.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RA   Favello T.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U39855; AAA81082.1; -.
KW   Lectin.
SQ   SEQUENCE   467 AA;  53169 MW;  7D9BBAB61830431B CRC64;


  Query Match             53.8%;  Score 49;  DB 5;  Length 467;
  Best Local Similarity   83.3%;  Pred. No. 31;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 WPWWPW 10
          |||| |
Db    201 WPWWVW 206


RESULT  14
Q26589
ID   Q26589      PRELIMINARY;      PRT;   528 AA.
AC   Q26589;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MYOSIN II HEAVY CHAIN (FRAGMENT).
OS   Schistosoma mansoni (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC   Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae;
OC   Schistosoma.
OX   NCBI_TaxID=6183;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PUERTO RICAN;
RA   Amory L.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PUERTO RICAN;
RX   MEDLINE=93056536; PubMed=1431131;
RA   Soisson L.M., Masterson C.P., Tom T.D., McNally M.T., Lowell G.H.,
RA   Strand M.;
RT   "Induction of protective immunity in mice using a 62-kDa recombinant
RT   fragment of a Schistosoma mansoni surface antigen.";
RL   J. Immunol. 149:3612-3620(1992).
DR   EMBL; X65591; CAA46548.1; -.
DR   HSSP; P08799; 1MMD.
DR   InterPro; IPR001637; GlnA_adenyltn.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR000533; Tropomyosin.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00194; TROPOMYOSIN.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   PROSITE; PS00182; GLNA_ADENYLATION; UNKNOWN_1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin.
FT   NON_TER       1      1
SQ   SEQUENCE   528 AA;  61622 MW;  AF075D13EB249B4C CRC64;


  Query Match             53.8%;  Score 49;  DB 5;  Length 528;
  Best Local Similarity   62.5%;  Pred. No. 35;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWW 8
            :|: ||||
Db    106 VLRNWPWW 113


RESULT  15
Q9TY57
ID   Q9TY57      PRELIMINARY;      PRT;   528 AA.
AC   Q9TY57;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MYOSIN HEAVY CHAIN (FRAGMENT).
OS   Schistosoma japonicum (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC   Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae;
OC   Schistosoma.
OX   NCBI_TaxID=6182;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CHINESE MAINLAND STRAIN;
RX   MEDLINE=99144454; PubMed=9990643;
RA   Zhang Y.B., Taylor M.G., Bickle Q.D.;
RT   "Schistosoma japonicum myosin: cloning, expression and vaccination
RT   studies with the homologue of the S.mansoni myosin fragment IrV-5.";
RL   Parasite Immunol. 20:583-594(1998).
DR   EMBL; U55313; AAC82221.1; -.
DR   HSSP; P08799; 1MMD.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR000533; Tropomyosin.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00194; TROPOMYOSIN.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
FT   NON_TER       1      1
FT   NON_TER     528    528
SQ   SEQUENCE   528 AA;  61406 MW;  C54A31F540F5EE05 CRC64;


  Query Match             53.8%;  Score 49;  DB 5;  Length 528;
  Best Local Similarity   62.5%;  Pred. No. 35;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ILKKWPWW 8
            :|: ||||
Db    106 VLRNWPWW 113


Search completed: January  4, 2002, 08:47:24
Job time: 412 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:47:47 ; Search time 18.1 Seconds
(without alignments)
26.334 Million cell updates/sec

Title: US-09-444-281-35
Perfect score: 91
Sequence: 1 ILKKWPWWPWRRK 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	73	80.2	144	1	INDC_BOVIN	P33046 bos taurus
2	54	59.3	1173	1	VGL2_CVH22	P15423 human coron
3	49	53.8	492	1	ADRO_BOVIN	P08165 bos taurus
4	49	53.8	715	1	YD55_MYCTU	Q11025 mycobacteri
5	47	51.6	253	1	Y945_MYCTU	P71564 mycobacteri
6	46.5	51.1	1154	1	VGL2_IBVD2	P12722 avian infec
7	46.5	51.1	1162	1	VGL2_IBVB	P11223 avian infec
8	46.5	51.1	1162	1	VGL2_IBVK	P12650 avian infec
9	46.5	51.1	1162	1	VGL2_IBVM	P12651 avian infec
10	46.5	51.1	1163	1	VGL2_IBV6	P05135 avian infec
11	46	50.5	196	1	YA05_SCHPO	Q09677 schizosacch
12	45	49.5	397	1	MML6_MYCTU	Q10773 mycobacteri
13	45	49.5	505	1	TRPE_PSESS	P21689 pseudomonas
14	45	49.5	512	1	FEN2_YEAST	P25621 saccharomyc
15	45	49.5	964	1	MML5_MYCTU	O53784 mycobacteri
16	45	49.5	967	1	MML4_MYCTU	O53735 mycobacteri
17	45	49.5	968	1	MML2_MYCTU	Q11171 mycobacteri
18	45	49.5	1108	1	CN3B_RAT	Q63085 rattus norv
19	45	49.5	1225	1	VGL2_CVPR8	P27655 porcine res
20	45	49.5	1225	1	VGL2_CVPRM	P24413 porcine res
21	45	49.5	1235	1	VGL2_CVMJH	P11225 murine coro
22	45	49.5	1324	1	VGL2_CVMA5	P11224 murine coro
23	45	49.5	1353	1	VGL2_CVHOC	P36334 human coron
24	45	49.5	1363	1	VGL2_CVBF	P25190 bovine coro
25	45	49.5	1363	1	VGL2_CVBL9	P25191 bovine coro
26	45	49.5	1363	1	VGL2_CVBLY	P25192 bovine coro
27	45	49.5	1363	1	VGL2_CVBM	P15777 bovine coro
28	45	49.5	1363	1	VGL2_CVBQ	P25193 bovine coro
29	45	49.5	1363	1	VGL2_CVBV	P25194 bovine coro
30	45	49.5	1376	1	VGL2_CVM4	P22432 murine coro
31	45	49.5	1376	1	VGL2_CVMJC	Q02385 murine coro
32	45	49.5	1447	1	VGL2_CVPPR	Q02167 porcine tra
33	45	49.5	1447	1	VGL2_CVPPU	P07946 porcine tra
34	45	49.5	1447	1	VGL2_CVPRT	Q01977 porcine tra
35	45	49.5	1449	1	VGL2_CVPFS	P18450 porcine tra
36	45	49.5	1449	1	VGL2_CVPMI	P33470 porcine tra
37	45	49.5	1451	1	VGL2_CVCAI	P36300 canine ente
38	45	49.5	1452	1	VGL2_FIPV	P10033 feline infe
39	45	49.5	2116	1	MYS2_DICDI	P08799 dictyosteli
40	44	48.4	53	1	YDH3_PLAFS	P14589 plasmodium
41	44	48.4	151	1	YBBJ_ECOLI	P75709 escherichia
42	44	48.4	361	1	FUT3_HUMAN	P21217 homo sapien
43	44	48.4	372	1	FUT3_PANTR	O19058 pan troglod
44	44	48.4	451	1	MENE_ECOLI	P37353 escherichia
45	44	48.4	535	1	YDW6_SCHPO	O13912 schizosacch

ALIGNMENTS

```
RESULT   1
INDC_BOVIN
ID   INDC_BOVIN     STANDARD;      PRT;   144 AA.
AC   P33046;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   INDOLICIDIN PRECURSOR.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=92392368; PubMed=1520337;
RA   del Sal G., Storici P., Schneider C., Romeo D., Zanetti M.;
RT   "cDNA cloning of the neutrophil bactericidal peptide indolicidin.";
RL   Biochem. Biophys. Res. Commun. 187:467-472(1992).
RN   [2]
RP   SEQUENCE OF 131-143.
RC   TISSUE=Neutrophils;
RX   MEDLINE=92165771; PubMed=1537821;
RA   Selsted M.E., Novotny M.J., Morris W.L., Tang Y.-Q., Smith W.,
RA   Cullor J.S.;
RT   "Indolicidin, a novel bactericidal tridecapeptide amide from
RT   neutrophils.";
RL   J. Biol. Chem. 267:4292-4295(1992).
CC   -!- FUNCTION: POTENT MICROBICIDAL ACTIVITY, ACTIVE AGAINST
CC       STAPHYLOCOCCUS AUREUS AND ESCHERICHIA COLI.
CC   -!- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.
CC   -!- PTM: ELASTASE MIGHT BE RESPONSIBLE FOR ITS MATURATION.
CC   -!- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67340; CAA47755.1; -.
DR   PIR; JC1222; JC1222.
DR   PIR; A42387; A42387.
DR   InterPro; IPR001894; Cathelicidin.
DR   Pfam; PF00666; Cathelicidins; 1.
DR   ProDom; PD001838; Cathelicidin; 1.
DR   PROSITE; PS00946; CATHELICIDINS_1; 1.
DR   PROSITE; PS00947; CATHELICIDINS_2; 1.
KW   Antibiotic; Amidation; Signal.
FT   SIGNAL       1     29       POTENTIAL.
FT   PROPEP      30    130
FT   PEPTIDE    131    143       INDOLICIDIN.
FT   MOD_RES     30     30       PYRROLIDONE CARBOXYLIC ACID (BY
```

```
FT                                SIMILARITY).
FT   DISULFID     85     96       BY SIMILARITY.
FT   DISULFID    107    124       BY SIMILARITY.
FT   MOD_RES     143    143       AMIDATION (G-144 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   144 AA;  16479 MW;  E3B1CBBE55C09911 CRC64;


  Query Match             80.2%;  Score 73;  DB 1;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 0.00094;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 KWPWWPWRR 12
           |||||||||
Db     135 KWPWWPWRR 143


RESULT   2
VGL2_CVH22
ID   VGL2_CVH22     STANDARD;      PRT;  1173 AA.
AC   P15423;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90264837; PubMed=2345367;
RA   Raabe T., Schelle-Prinz B., Siddell S.G.;
RT   "Nucleotide sequence of the gene encoding the spike glycoprotein of
RT   human coronavirus HCV 229E.";
RL   J. Gen. Virol. 71:1065-1073(1990).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC       AND IN SYNCYTIUM FORMATION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X16816; CAA34723.1; -.
DR   PIR; A34766; VGIHHC.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     15
FT   CHAIN        16   1173       E2 GLYCOPROTEIN.
FT   DOMAIN       16   1115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1116   1135       POTENTIAL.
FT   DOMAIN     1136   1173       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1136   1157       CYS-RICH.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     62     62       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    171    171       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    440    440       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    464    464       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    518    518       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    538    538       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    581    581       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    587    587       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    671    671       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    930    930       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1015   1015       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1020   1020       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1037   1037       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1066   1066       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1076   1076       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1082   1082       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1096   1096       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1173 AA;  128639 MW;  B9CA9A41A796B3BD CRC64;


  Query Match             59.3%;  Score 54;  DB 1;  Length 1173;
  Best Local Similarity   85.7%;  Pred. No. 2.5;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 KWPWWPW 10
           ||||| |
Db    1113 KWPWWVW 1119


RESULT   3
ADRO_BOVIN
ID   ADRO_BOVIN     STANDARD;      PRT;   492 AA.
AC   P08165;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NADPH:ADRENODOXIN OXIDOREDUCTASE, MITOCHONDRIAL PRECURSOR
DE   (EC 1.18.1.2) (ADRENODOXIN REDUCTASE) (AR) (FERREDOXIN-NADP(+)
DE   REDUCTASE).
GN   FDXR OR ADXR.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=94177140; PubMed=8130767;
RA   Takata Y., Sagara Y., Kono A., Sekimizu K., Horiuchi T.;
RT   "Gene structure of bovine adrenodoxin reductase.";
RL   Biol. Pharm. Bull. 16:1200-1206(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=88198050; PubMed=3448086;
RA   Sagara Y., Takata Y., Miyata T., Hara T., Horiuchi T.;
RT   "Cloning and sequence analysis of adrenodoxin reductase cDNA from
RT   bovine adrenal cortex.";
RL   J. Biochem. 102:1333-1336(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87270696; PubMed=3038094;
RA   Nonaka Y., Murakami H., Yabusaki Y., Kuramitsu S., Kagamiyama H.,
RA   Yamano T., Okamoto M.;
RT   "Molecular cloning and sequence analysis of full-length cDNA for mRNA
RT   of adrenodoxin oxidoreductase from bovine adrenal cortex.";
RL   Biochem. Biophys. Res. Commun. 145:1239-1247(1987).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal cortex;
RX   MEDLINE=89170752; PubMed=2924777;
RA   Hanukoglu I., Gutfinger T.;
```

```
RT   "cDNA sequence of adrenodoxin reductase. Identification of NADP-
RT   binding sites in oxidoreductases.";
RL   Eur. J. Biochem. 180:479-484(1989).
RN   [5]
RP   SEQUENCE OF N-TERMINUS, AND PARTIAL SEQUENCE.
RC   TISSUE=Adrenal cortex;
RX   MEDLINE=88082777; PubMed=3691502;
RA   Hanukoglu I., Gutfinger T., Haniu M., Shively J.E.;
RT   "Isolation of a cDNA for adrenodoxin reductase (ferredoxin-NADP+
RT   reductase). Implications for mitochondrial cytochrome P-450 systems.";
RL   Eur. J. Biochem. 169:449-455(1987).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 33-492.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=99299392; PubMed=10369776;
RA   Ziegler G.A., Vonrhein C., Hanukoglu I., Schulz G.E.;
RT   "The structure of adrenodoxin reductase of mitochondrial P450 systems:
RT   electron transfer for steroid biosynthesis.";
RL   J. Mol. Biol. 289:981-990(1999).
CC   -!- FUNCTION: SERVES AS THE FIRST ELECTRON TRANSFER PROTEIN IN ALL THE
CC       MITOCHONDRIAL P450 SYSTEMS. INCLUDING CHOLESTEROL SIDE CHAIN
CC       CLEAVAGE IN ALL STEROIDOGENIC TISSUES, STEROID 11-BETA
CC       HYDROXYLATION IN THE ADRENAL CORTEX, 25-OH-VITAMIN D3-24
CC       HYDROXYLATION IN THE KIDNEY, AND STEROL C-27 HYDROXYLATION IN THE
CC       LIVER.
CC   -!- CATALYTIC ACTIVITY: REDUCED ADRENODOXIN + NADP(+) = OXIDIZED
CC       ADRENODOXIN + NADPH.
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: CHOLESTEROL SIDE-CHAIN-CLEAVAGE SYSTEM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM (SHOWN HERE) AND A
CC       LONG FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE LONG FORM
CC       REPRESENTS 10-20% OF ALL ADRENODOXIN REDUCTASE MRNA. AND SEEMS TO
CC       BE INACTIVE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17029; AAA30362.1; -.
DR   EMBL; D00211; BAA00150.1; -.
DR   EMBL; X13736; CAA32002.1; -.
DR   PIR; A29604; A29604.
DR   PIR; JS0390; JS0390.
DR   PIR; S03558; S03558.
DR   PIR; JT0751; JT0751.
DR   PDB; 1CJC; 12-APR-99.
DR   PDB; 1E1L; 02-JUN-00.
DR   InterPro; IPR000759; Adrndx_redctse.
DR   PRINTS; PR00419; ADXRDTASE.
KW   Electron transport; Oxidoreductase; Flavoprotein; NADP; FAD;
KW   Mitochondrion; Transit peptide; Alternative splicing; 3D-structure.
FT   TRANSIT       1     32       MITOCHONDRION.
FT   CHAIN        33    492       NADPH:ADRENODOXIN OXIDOREDUCTASE.
FT   VARSPLIC    204    204       E -> EVLLLCQ (IN LONG ISOFORM).
FT   CONFLICT     77     77       G -> R (IN REF. 3).
FT   CONFLICT     81     94       FGVAPDHPEVKNVI -> VWLALTTPRSRMLL (IN REF.
FT                                3).
FT   CONFLICT    124    128       QDAYH -> RVYRLT (IN REF. 3).
FT   CONFLICT    268    268       K -> R (IN REF. 3).
FT   CONFLICT    317    318       PS -> RL (IN REF. 3).
FT   CONFLICT    323    333       RAAGIRLAVTR -> ARRSAWQSPE (IN REF. 3).
FT   CONFLICT    341    352       TRAVPTGDVEDL -> HPGSAHWGCGGP (IN REF. 3).
SQ   SEQUENCE   492 AA;  54338 MW;  E68F6F5D18F53131 CRC64;


  Query Match             53.8%;  Score 49;  DB 1;  Length 492;
  Best Local Similarity   83.3%;  Pred. No. 5.4;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 WPWWPW 10
           | ||||
Db       6 WRWWPW 11


RESULT   4
YD55_MYCTU
ID   YD55_MYCTU     STANDARD;      PRT;   715 AA.
AC   Q11025;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 78.2 KDA PROTEIN RV1355C.
GN   RV1355C OR MT1398 OR MTCY02B10.19C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z75555; CAA99988.1; -.
DR   EMBL; AE007012; AAK45661.1; ALT_INIT.
DR   TIGR; MT1398; -.
DR   TubercuList; Rv1355c; -.
DR   InterPro; IPR000594; ThiF_family.
DR   Pfam; PF00899; ThiF_family; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   715 AA;  78181 MW;  455495248A56041C CRC64;


  Query Match             53.8%;  Score 49;  DB 1;  Length 715;
  Best Local Similarity   60.0%;  Pred. No. 7.6;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 KKWPWWPWRR 12
           |:|  ::||||
Db      64 KRWAYYPWRR 73
```

```
RESULT   5
Y945_MYCTU
ID   Y945_MYCTU     STANDARD;      PRT;   253 AA.
AC   P71564;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE OXIDOREDUCTASE RV0945 (EC 1.-.-.-).
GN   RV0945 OR MT0971 OR MTCY10D7.29C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z79700; CAB02005.1; -.
DR   EMBL; AE006982; AAK45219.1; -.
DR   TIGR; MT0971; -.
DR   TubercuList; Rv0945; -.
DR   InterPro; IPR002198; ADH_short.
DR   Pfam; PF00106; adh_short; 1.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Hypothetical protein; Oxidoreductase; Complete proteome.
FT   ACT_SITE    159    159       BY SIMILARITY.
SQ   SEQUENCE   253 AA;  27138 MW;  BAD937208842DA12 CRC64;


  Query Match             51.6%;  Score 47;  DB 1;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 PWWPW 10
           |||||
Db     230 PWWPW 234


RESULT   6
```

```
VGL2_IBVD2
ID   VGL2_IBVD2     STANDARD;      PRT;  1154 AA.
AC   P12722; Q66176; Q66177;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE   [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN   S.
OS   Avian infectious bronchitis virus (strain D274) (IBV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11124;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89386000; PubMed=2550899;
RA   Jordi B.J.A.M., Kremers D.A.W.M., Kusters H.G., van der Zeijst B.A.M.;
RT   "Nucleotide sequence of the gene coding for the peplomer protein (=
RT   spike protein) of infectious bronchitis virus, strain D274.";
RL   Nucleic Acids Res. 17:6726-6726(1989).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15832; CAA33837.1; -.
DR   PIR; A34300; VGIHIB.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19   1154       E2 GLYCOPROTEIN.
FT   CHAIN        19    538       SPIKE PROTEIN S1.
FT   CHAIN       539   1154       SPIKE PROTEIN S2.
FT   DOMAIN     1121   1138       CYS-RICH.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    102    102       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    179    179       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    213    213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    265    265       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    307    307       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    514    514       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    531    531       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    592    592       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    948    948       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    961    961       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    980    980       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1015   1015       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1039   1039       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1052   1052       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1075   1075       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
SQ   SEQUENCE   1154 AA;  127502 MW;  D79F37AF89F1A37F CRC64;


  Query Match             51.1%;  Score 46.5;  DB 1;  Length 1154;
  Best Local Similarity   61.5%;  Pred. No. 26;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
             |||   |||| : |
Db    1086 ILKTYIKWPWYVW 1098


RESULT   7
VGL2_IBVB
ID   VGL2_IBVB      STANDARD;      PRT;  1162 AA.
AC   P11223; P05134;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE   [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN   S.
OS   Avian infectious bronchitis virus (strain Beaudette) (IBV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11122;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85159540; PubMed=2984314;
RA   Binns M.M., Boursnell M.E.G., Cavanagh D., Pappind D.J.C.,
RA   Brown T.D.K.;
RT   "Cloning and sequencing of the gene encoding the spike protein of the
RT   coronavirus IBV.";
RL   J. Gen. Virol. 66:719-726(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87085499; PubMed=3025348;
RA   Binns M.M., Boursnell M.E.G., Tomley F.M., Brown D.K.;
RT   "Comparison of the spike precursor sequences of coronavirus IBV
RT   strains M41 and 6/82 with that of IBV Beaudette.";
RL   J. Gen. Virol. 67:2825-2831(1986).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95169; AAA70235.1; -.
DR   EMBL; X02342; CAA26201.1; -.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19   1162       E2 GLYCOPROTEIN.
FT   CHAIN        19    537       SPIKE PROTEIN S1.
FT   CHAIN       538   1162       SPIKE PROTEIN S2.
FT   DOMAIN     1120   1137       CYS-RICH.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     77     77       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    247    247       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    276    276       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    447    447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    669    669       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    714    714       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    960    960       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    979    979       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1014   1014       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1038   1038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1051   1051       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1074   1074       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1162 AA;  128046 MW;  0BAAD58113C8EBD5 CRC64;


  Query Match             51.1%;  Score 46.5;  DB 1;  Length 1162;
  Best Local Similarity   61.5%;  Pred. No. 26;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
             |||   |||| : |
Db    1085 ILKTYIKWPWYVW 1097


RESULT   8
VGL2_IBVK
ID   VGL2_IBVK      STANDARD;      PRT;  1162 AA.
AC   P12650;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE   [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN   S.
OS   Avian infectious bronchitis virus (strain KB8523) (IBV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11126;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88306251; PubMed=2841803;
RA   Sutou S., Sato S., Okabe T., Nakai M., Sasaki N.;
RT   "Cloning and sequencing of genes encoding structural proteins of
RT   avian infectious bronchitis virus.";
RL   Virology 165:589-595(1988).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21515; AAA66578.1; -.
DR   PIR; B29249; VGIHAK.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19   1162       E2 GLYCOPROTEIN.
FT   CHAIN        19    537       SPIKE PROTEIN S1.
```

```
FT   CHAIN       538   1162       SPIKE PROTEIN S2.
FT   DOMAIN     1120   1137       CYS-RICH.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     77     77       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    247    247       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    276    276       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    447    447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    669    669       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    714    714       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    960    960       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    979    979       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1014   1014       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1051   1051       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1058   1058       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1074   1074       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1162 AA;  128537 MW;  2299036B3597EA8F CRC64;

  Query Match           51.1%;  Score 46.5;  DB 1;  Length 1162;
  Best Local Similarity   61.5%;  Pred. No. 26;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
           |||   |||| : |
Db    1085 ILKTYIKWPWYVW 1097


RESULT   9
VGL2_IBVM
ID   VGL2_IBVM      STANDARD;      PRT;  1162 AA.
AC   P12651;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE   [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN   S.
OS   Avian infectious bronchitis virus (strain M41) (IBV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11127;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87021475; PubMed=2429473;
RA   Niesters H.G.M., Lenstra J.A., Spaan W.J.M., Zijderveld A.J.,
RA   Bleumink-Pluym N.M.C., Hong F., van Scharrenburg G.J.M.,
RA   Horzinek M.C., van der Zeijst B.A.M.;
RT   "The peplomer protein sequence of the M41 strain of coronavirus IBV
RT   and its comparison with Beaudette strains.";
RL   Virus Res. 5:253-263(1986).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21883; AAA66575.1; -.
DR   EMBL; A24863; CAA01736.1; -.
DR   PIR; S07421; S07421.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19   1162       E2 GLYCOPROTEIN.
FT   CHAIN        19    537       SPIKE PROTEIN S1.
FT   CHAIN       538   1162       SPIKE PROTEIN S2.
FT   DOMAIN     1120   1137       CYS-RICH.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     77     77       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    247    247       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    276    276       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    447    447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    669    669       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    714    714       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    960    960       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    979    979       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1014   1014       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1038   1038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1051   1051       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1074   1074       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1162 AA;  128077 MW;  3C9CC70938492DDA CRC64;

  Query Match           51.1%;  Score 46.5;  DB 1;  Length 1162;
  Best Local Similarity   61.5%;  Pred. No. 26;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
           |||   |||| : |
Db    1085 ILKTYIKWPWYVW 1097


RESULT  10
VGL2_IBV6
ID   VGL2_IBV6      STANDARD;      PRT;  1163 AA.
AC   P05135;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE   [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN   S.
OS   Avian infectious bronchitis virus (strain 6/82) (IBV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11121;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87085499; PubMed=3025348;
RA   Binns M.M., Boursnell M.E.G., Tomley F.M., Brown D.K.;
RT   "Comparison of the spike precursor sequences of coronavirus IBV
RT   strains M41 and 6/82 with that of IBV Beaudette.";
RL   J. Gen. Virol. 67:2825-2831(1986).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04723; CAA28432.1; -.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     18
FT   CHAIN        19   1163       E2 GLYCOPROTEIN.
FT   CHAIN        19    538       SPIKE PROTEIN S1.
FT   CHAIN       539   1163       SPIKE PROTEIN S2.
FT   DOMAIN     1121   1138       CYS-RICH.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    102    102       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    179    179       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    213    213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    265    265       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    307    307       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    514    514       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    531    531       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    592    592       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    948    948       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    961    961       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    980    980       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1015   1015       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1039   1039       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1052   1052       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1075   1075       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1163 AA;  128684 MW;  8FE344CF2995478C CRC64;


  Query Match             51.1%;  Score 46.5;  DB 1;  Length 1163;
  Best Local Similarity   61.5%;  Pred. No. 26;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
           |||   |||| : |
Db    1086 ILKTYIKWPWYVW 1098


RESULT  11
YA05_SCHPO
ID   YA05_SCHPO     STANDARD;      PRT;   196 AA.
AC   Q09677;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL 22.1 KDA PROTEIN C5H10.05C IN CHROMOSOME I.
GN   SPAC5H10.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO BACTERIAL MODULATOR OF DRUG ACTIVITY B
CC       (MDAB).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49811; CAA89955.1; -.
DR   InterPro; IPR003680; NADHdh_2.
DR   Pfam; PF02525; NADHdh_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   196 AA;  22104 MW;  436764DA9E26074C CRC64;


  Query Match             50.5%;  Score 46;  DB 1;  Length 196;
  Best Local Similarity   47.1%;  Pred. No. 5.9;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    4;  Gaps    2;

Qy       1 ILKKWP-WW---PWRRK 13
             |: :|| ||   ||: |
Db      62 IIYQWPGWWMGTPWKLK 78


RESULT  12
MML6_MYCTU
ID   MML6_MYCTU     STANDARD;      PRT;   397 AA.
AC   Q10773;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN MMPL6.
GN   MMPL6 OR RV1557 OR MT1608 OR MTCY48.08C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74020; CAA98334.1; -.
DR   EMBL; AE007027; AAK45875.1; -.
DR   TIGR; MT1608; -.
DR   TubercuList; Rv1557; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM    161    181       POTENTIAL.
FT   TRANSMEM    190    210       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    242    262       POTENTIAL.
FT   TRANSMEM    293    313       POTENTIAL.
FT   TRANSMEM    330    350       POTENTIAL.
SQ   SEQUENCE   397 AA;  42421 MW;  678DC86E24472BF4 CRC64;


  Query Match             49.5%;  Score 45;  DB 1;  Length 397;
  Best Local Similarity   54.5%;  Pred. No. 15;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWPWR 11
           :| :| ||| |
Db     348 LLGRWFWWPQR 358


RESULT  13
TRPE_PSESS
ID   TRPE_PSESS     STANDARD;      PRT;   505 AA.
AC   P21689;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN   TRPE.
OS   Pseudomonas syringae (pv. savastanoi).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=29438;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91100331; PubMed=1987141;
RA   da Costa E., Silva O., Kosuge T.;
RT   "Molecular characterization and expression analysis of the
RT   anthranilate synthase gene of Pseudomonas syringae subsp.
RT   savastanoi.";
RL   J. Bacteriol. 173:463-471(1991).
CC   -!- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
CC       PYRUVATE + L-GLUTAMATE.
CC   -!- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC   -!- SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II (BY
CC       SIMILARITY).
CC   -!- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
CC       USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
CC       GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55911; AAA26016.1; -.
DR   PIR; A39128; A39128.
DR   InterPro; IPR000350; Chorismate_bind.
DR   Pfam; PF00425; chorismate_bind; 1.
DR   PRINTS; PR00095; ANTSNTHASEI.
DR   ProDom; PD000779; Chorismate_bind; 1.
KW   Tryptophan biosynthesis; Lyase.
SQ   SEQUENCE   505 AA;  56084 MW;  A38E81931331F6BB CRC64;


  Query Match             49.5%;  Score 45;  DB 1;  Length 505;
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7 WWPWRRK 13
           |||| |:
Db     485 WWPWPRR 491


RESULT  14
FEN2_YEAST
ID   FEN2_YEAST     STANDARD;      PRT;   512 AA.
AC   P25621;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PROBABLE TRANSPORTER FEN2.
GN   FEN2 OR YCR028C OR YCR28C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cederberg H., Hohmann S., Schaaff-Gerstenschlaeger I., Huse K.,
RA   Zimmermann F.K.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93070619; PubMed=1332309;
RA   Carbone M.L.A., Panzeri L., Falconi M.M., Carcano C., Plevani P.,
RA   Lucchini G.;
RT   "Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III
RT   from strain AB972: evidence for a Ty insertion and functional
RT   analysis of open reading frame YCR28.";
RL   Yeast 8:805-812(1992).
RN   [3]
RP   SIMILARITY TO DAL5 FAMILY.
RX   MEDLINE=94147996; PubMed=8313894;
RA   Koonin E.V., Bork P., Sander C.;
RT   "Yeast chromosome III: new gene functions.";
RL   EMBO J. 13:493-503(1994).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=96367594; PubMed=8771708;
RA   Marcireau C., Joets J., Pousset D., Guilloton M., Karst F.;
RT   "FEN2: a gene implicated in the catabolite repression-mediated
RT   regulation of ergosterol biosynthesis in yeast.";
RL   Yeast 12:531-539(1996).
CC   -!- FUNCTION: INVOLVED IN THE CATABOLITE REPRESSION-MEDIATED
CC       REGULATION OF ERGOSTEROL BIOSYNTHESIS AND IN FENPROPIMORPH
CC       RESISTANCE.
```

```
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ALLANTOATE PERMEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAA42320.1; -.
DR   PIR; S19439; S19439.
DR   PIR; S25336; S25336.
DR   SGD; S0000623; FEN2.
KW   Transmembrane; Transport.
FT   TRANSMEM     28     48       POTENTIAL.
FT   TRANSMEM     80    100       POTENTIAL.
FT   TRANSMEM    103    123       POTENTIAL.
FT   TRANSMEM    133    153       POTENTIAL.
FT   TRANSMEM    165    185       POTENTIAL.
FT   TRANSMEM    199    219       POTENTIAL.
FT   TRANSMEM    272    292       POTENTIAL.
FT   TRANSMEM    313    333       POTENTIAL.
FT   TRANSMEM    343    363       POTENTIAL.
FT   TRANSMEM    373    393       POTENTIAL.
FT   TRANSMEM    402    422       POTENTIAL.
FT   TRANSMEM    435    455       POTENTIAL.
FT   CONFLICT    104    104       W -> V (IN REF. 2).
SQ   SEQUENCE   512 AA;  58256 MW;  361942E74C62B3B4 CRC64;


  Query Match             49.5%;  Score 45;  DB 1;  Length 512;
  Best Local Similarity   62.5%;  Pred. No. 20;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWW 8
           :||:| ||
Db     268 VLKRWHWW 275


RESULT  15
MML5_MYCTU
ID   MML5_MYCTU     STANDARD;      PRT;   964 AA.
AC   O53784;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN MMPL5.
GN   MMPL5 OR RV0676C OR MT0705 OR MTV040.04C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021943; CAA17459.1; -.
DR   EMBL; AE006964; AAK44930.1; -.
DR   TIGR; MT0705; -.
DR   TubercuList; Rv0676c; -.
DR   InterPro; IPR001036; ACR_tran.
DR   PRINTS; PR00702; ACRIFLAVINRP.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     31     51       POTENTIAL.
FT   TRANSMEM    203    223       POTENTIAL.
FT   TRANSMEM    230    250       POTENTIAL.
FT   TRANSMEM    255    275       POTENTIAL.
FT   TRANSMEM    302    322       POTENTIAL.
FT   TRANSMEM    340    360       POTENTIAL.
FT   TRANSMEM    389    409       POTENTIAL.
FT   TRANSMEM    745    765       POTENTIAL.
FT   TRANSMEM    774    794       POTENTIAL.
FT   TRANSMEM    803    823       POTENTIAL.
FT   TRANSMEM    826    846       POTENTIAL.
FT   TRANSMEM    880    900       POTENTIAL.
FT   TRANSMEM    901    921       POTENTIAL.
FT   CONFLICT    948    948       I -> V (IN REF. 2).
SQ   SEQUENCE   964 AA;  104784 MW;  B7C945940A1176BD CRC64;


  Query Match             49.5%;  Score 45;  DB 1;  Length 964;
  Best Local Similarity   66.7%;  Pred. No. 35;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWWP 9
           :| || |||
Db     932 LLGKWFWWP 940


Search completed: January  4, 2002, 08:47:48
Job time: 406 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:41:31 ; Search time 27.18 Seconds
(without alignments)
36.434 Million cell updates/sec

Title: US-09-444-281-35
Perfect score: 91
Sequence: 1 ILKKWPWWPWRRK 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	73	80.2	144	1	JC1222	indolicidin precur
2	54	59.3	1173	1	VGIHHC	E2 glycoprotein pr
3	51	56.0	299	2	T12505	hypothetical prote
4	49	53.8	298	2	B72492	hypothetical prote
5	49	53.8	498	1	JT0751	ferredoxin--NADP+
6	49	53.8	527	2	S33068	myosin heavy chain
7	49	53.8	715	2	B70741	probable moeY prot
8	49	53.8	1940	2	A59287	myosin heavy chain
9	48	52.7	111	2	T29295	hypothetical prote
10	47.5	52.2	114	2	T36208	hypothetical prote
11	47	51.6	248	2	S23449	NADH oxidase (H2O2
12	47	51.6	253	2	G70715	hypothetical prote
13	46.5	51.1	352	2	S77354	histidinol-phospha
14	46.5	51.1	621	2	S37664	peplomeric polypro
15	46.5	51.1	630	2	S37663	peplomeric polypro
16	46.5	51.1	1154	1	VGIHIB	E2 glycoprotein pr
17	46.5	51.1	1162	1	VGIHAK	E2 glycoprotein pr
18	46.5	51.1	1162	2	S07421	E2 glycoprotein pr
19	46.5	51.1	1162	2	S14939	E2 glycoprotein pr
20	46.5	51.1	1162	2	S14940	E2 glycoprotein pr
21	46	50.5	196	2	S55483	modulator of drug
22	46	50.5	617	2	T22175	hypothetical prote
23	46	50.5	623	2	T22177	hypothetical prote
24	46	50.5	1333	2	S65812	RNA-directed DNA p
25	45	49.5	273	2	T82646	monofunctional bio
26	45	49.5	276	2	B83161	probable short-cha
27	45	49.5	397	2	B70763	probable membrane
28	45	49.5	448	2	H72376	hypothetical prote
29	45	49.5	505	2	A39128	anthranilate synth
30	45	49.5	512	2	S19439	probable membrane
31	45	49.5	964	2	E70826	probable membrane
32	45	49.5	967	2	C70831	probable mmpL4 pro
33	45	49.5	968	2	F70746	probable mmpL2 pro
34	45	49.5	1108	2	A48508	cyclic-nucleotide
35	45	49.5	1225	1	S24284	E2 glycoprotein pr
36	45	49.5	1225	2	A36607	E2 glycoprotein -
37	45	49.5	1235	1	VGIHMJ	E2 glycoprotein pr
38	45	49.5	1324	1	VGIH59	E2 glycoprotein pr
39	45	49.5	1353	1	JQ2168	E2 glycoprotein pr
40	45	49.5	1361	2	S29998	surface protein -
41	45	49.5	1362	2	A37474	surface glycoprote
42	45	49.5	1363	1	VGIHNM	E2 glycoprotein pr
43	45	49.5	1363	1	VGIHQU	E2 glycoprotein pr
44	45	49.5	1363	1	VGIHF1	E2 glycoprotein pr
45	45	49.5	1363	1	VGIHL9	E2 glycoprotein pr

ALIGNMENTS

RESULT 1
JC1222
indolicidin precursor - bovine
N;Alternate names: antimicrobial peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JC1222; A42387; S25664
R;del Sal, G.; Storici, P.; Schneider, C.; Romeo, D.; Zanetti, M.
Biochem. Biophys. Res. Commun. 187, 467-472, 1992
A;Title: cDNA cloning of the neutrophil bactericidal peptide indolicidin.
A;Reference number: JC1222; MUID:92392368
A;Accession: JC1222
A;Molecule type: mRNA
A;Residues: 1-144 <SAL>
A;Cross-references: EMBL:X67340; NID:g462; PIDN:CAA47755.1; PID:g463
A;Experimental source: bone marrow
R;Selsted, M.E.; Novotny, M.J.; Morris, W.L.; Tang, Y.Q.; Smith, W.; Cullor, J.S.
J. Biol. Chem. 267, 4292-4295, 1992
A;Title: Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils.
A;Reference number: A42387; MUID:92165771
A;Accession: A42387
A;Molecule type: protein
A;Residues: 131-143 <SEL>
A;Experimental source: neutrophils
A;Note: sequence extracted from NCBI backbone (NCBIP:83840)
C;Superfamily: cathelin; cystatin homology
C;Keywords: amidated carboxyl end
F;1-29/Domain: signal sequence #status predicted <SIG>
F;22-129/Domain: cystatin homology <CYS>
F;30-130/Domain: propeptide #status predicted <PRO>
F;131-143/Product: indolicidin #status experimental <MAT>
F;143/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 80.2%; Score 73; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      4 KWPWWPWRR 12
          |||||||||
Db    135 KWPWWPWRR 143
```

RESULT 2
VGIHHC
E2 glycoprotein precursor - human coronavirus (strain 229E)
N;Alternate names: peplomer glycoprotein; S glycoprotein; spike glycoprotein
C;Species: human coronavirus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C;Accession: A34766; S05460
R;Raabe, T.; Schelle-Prinz, B.; Siddell, S.G.

J. Gen. Virol. 71, 1065-1073, 1990
A;Title: Nucleotide sequence of the gene encoding the spike glycoprotein of human corona
A;Reference number: A34766; MUID:90264837
A;Accession: A34766
A;Molecule type: mRNA
A;Residues: 1-1173 <RAA>
A;Cross-references: EMBL:X16816; NID:g58926; PIDN:CAA34723.1; PID:g58927
A;Experimental source: strain 229E
R;Raabe, T.; Siddell, S.
Nucleic Acids Res. 17, 6387, 1989
A;Title: Nucleotide sequence of the human coronavirus HCV 229E mRNA 4 and mRNA 5 unique
A;Reference number: A34038; MUID:89366667
A;Accession: S05460
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1159-1173 <RA2>
A;Cross-references: EMBL:X15654; NID:g58921; PIDN:CAA33680.1; PID:g1334827
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-1173/Product: E2 glycoprotein #status predicted <MAT>
F;1116-1138/Domain: transmembrane #status predicted <TMN>
F;23,62,98,147,171,176,220,243,326,333,440,464,518,538,542,568,581,587,663,671,930,1015,

  Query Match 59.3%; Score 54; DB 1; Length 1173;
  Best Local Similarity 85.7%; Pred. No. 7;
  Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 KWPWWPW 10
 ||||| |
Db 1113 KWPWWVW 1119

RESULT 3
T12505
hypothetical protein DKFZp434C192.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C;Accession: T12505
R;Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z17527
A;Accession: T12505
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-299 <ANS>
A;Cross-references: EMBL:AL096753
A;Experimental source: adult testis; clone DKFZp434C192
C;Genetics:
A;Note: DKFZp434C192.1

  Query Match 56.0%; Score 51; DB 2; Length 299;
  Best Local Similarity 85.7%; Pred. No. 4.7;
  Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 PWWPWRR 12
 ||||| |
Db 37 PWWPWSR 43

RESULT 4
B72492
hypothetical protein APE2577 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: B72492
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: B72492
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-298 <KAW>
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81594.1; PID:d1045380; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE2577

  Query Match 53.8%; Score 49; DB 2; Length 298;
  Best Local Similarity 60.0%; Pred. No. 8.8;
  Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LKKWPWWPWR 11
 :|: ||| ||
Db 102 IKETPWWRWR 111

RESULT 5
JT0751
ferredoxin--NADP+ reductase (EC 1.18.1.2), long form precursor - bovine
N;Alternate names: adrenodoxin reductase
C;Species: Bos primigenius taurus (cattle)
C;Date: 14-Jul-1994 #sequence_revision 18-Oct-1996 #text_change 16-Jun-2000
C;Accession: JT0751; JT0079; JS0390; S03558; PS0003; A29604; S52100
R;Takata, Y.; Sagara, Y.; Kono, A.; Sekimizu, K.; Horiuchi, T.
Biol. Pharm. Bull. 16, 1200-1206, 1993
A;Title: Gene structure of bovine adrenodoxin reductase.
A;Reference number: JT0751; MUID:94177140
A;Accession: JT0751
A;Molecule type: DNA
A;Residues: 1-498 <TAK>
A;Cross-references: GB:D83475; NID:g1199916; PIDN:BAA11921.1; PID:g4521308
A;Experimental source: adrenal cortex
A;Note: the authors translated the codon GTC for residue 205 as Gly
R;Sagara, Y.; Takata, Y.; Miyata, T.; Hara, T.; Horiuchi, T.
J. Biochem. 102, 1333-1336, 1987
A;Title: Cloning and sequence analysis of adrenodoxin reductase cDNA from bovine adre
A;Reference number: JT0079; MUID:88198050
A;Accession: JT0079
A;Molecule type: mRNA
A;Residues: 1-204,211-498 <SAG>
A;Cross-references: GB:D00211; NID:g217433; PIDN:BAA00150.1; PID:g217434
A;Note: the deduced sequence is partially confirmed by amino acid sequencing of 15 is
R;Sagara, Y.
submitted to DDBJ, September 1989
A;Reference number: JS0390
A;Contents: revision, insertion of residues 205-210
A;Accession: JS0390
A;Molecule type: mRNA
A;Residues: 56-498 <SA2>
R;Hanukoglu, I.; Gutfinger, T.
Eur. J. Biochem. 180, 479-484, 1989
A;Title: cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites
A;Reference number: S03558; MUID:89170752
A;Accession: S03558
A;Molecule type: mRNA
A;Residues: 155-204,211-498 <HAN>
A;Cross-references: EMBL:X13736; NID:g65; PIDN:CAA32002.1; PID:g833776
A;Note: 405-Ser was also found
R;Hamamoto, I.; Kurokohchi, K.; Tanaka, S.; Ichikawa, Y.
Biochim. Biophys. Acta 953, 207-213, 1988
A;Title: Adrenoferredoxin-binding peptide of NADPH-adrenoferredoxin reductase.
A;Reference number: PS0003; MUID:88184054
A;Accession: PS0003
A;Molecule type: protein
A;Residues: 33-41,'S',43-62;260-283,'TM';496-498 <HAM>
A;Note: a cyanogen bromide peptide binds to adrenoferredoxin
R;Nonaka, Y.; Murakami, H.; Yabusaki, Y.; Kuramitsu, S.; Kagamiyama, H.; Yamano, T.;
Biochem. Biophys. Res. Commun. 145, 1239-1247, 1987
A;Title: Molecular cloning and sequence analysis of full-length cDNA for mRNA of adre
A;Reference number: A29604; MUID:87270696

```
A;Accession: A29604
A;Molecule type: mRNA
A;Residues: 1-76,'R',78-80,'VWLALTTPRSRMLL',95-123,'RVYRLT',129-204,211-273,'R',275-322,
A;Cross-references: GB:M17029; NID:g162628; PIDN:AAA30362.1; PID:g162629
A;Experimental source: adrenal cortex
R;Warburton, R.J.; Seybert, D.W.
Biochim. Biophys. Acta 1246, 39-46, 1995
A;Title: Structural and functional characterization of bovine adrenodoxin reductase by l
A;Reference number: S52100; MUID:95110846
A;Accession: S52100
A;Status: preliminary
A;Molecule type: protein
A;Residues: 'X',34-41,'X',43-48,'X',50-51;304-306,'X',308-309,'X',311-326 <WAR>
C;Comment: Ferredoxin--NADP+ reductase is localized in the matrix of adrenal cortex mitc
ferredoxin--NADP+ reductase, adrenodoxin and two forms of cytochrome P-450.
C;Genetics:
A;Introns: 27/1; 59/3; 91/3; 132/3; 170/3; 204/3; 246/3; 275/1; 341/3; 399/1; 456/1
C;Function:
A;Description: catalyzes the reversible reduction of NADP+ by reduced ferredoxin or redu
C;Superfamily: human ferredoxin--NADP+ reductase
C;Keywords: alternative splicing; flavoprotein; mitochondrion; monomer; NADP; oxidoreduc
F;1-32/Domain: transit peptide (mitochondrion) #status predicted <SIG>
F;33-498/Product: ferredoxin--NADP+ reductase, long form #status predicted <MAT>
F;33-204,211-498/Product: ferredoxin--NADP+ reductase, short form #status experimental <
F;40-70/Region: beta-alpha-beta FAD nucleotide-binding fold
F;180-190/Region: NADP binding #status predicted
F;281/Binding site: substrate (Lys) #status experimental


  Query Match             53.8%;  Score 49;  DB 1;  Length 498;
  Best Local Similarity   83.3%;  Pred. No. 14;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 WPWWPW 10
           | ||||
Db       6 WRWWPW 11


RESULT   6
S33068
myosin heavy chain - fluke (Schistosoma mansoni) (fragment)
N;Alternate names: surface antigen, 200K
C;Species: Schistosoma mansoni
C;Date: 22-Nov-1993 #sequence_revision 06-Sep-1996 #text_change 13-Feb-1998
C;Accession: S33068
R;Soisson, L.M.A.; Masterson, C.P.; Tom, T.D.; McNally, M.T.; Lowell, G.H.; Strand, M.
J. Immunol. 149, 3612-3620, 1992
A;Title: Induction of protective immunity in mice using a 62-kDa recombinant fragment of
A;Reference number: A46514; MUID:93056536
A;Accession: S33068
A;Molecule type: mRNA
A;Residues: 1-527 <SOI>
A;Cross-references: EMBL:X65591
A;Note: the authors translated the codon CAA for residue 346 as Lys
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: ATP; surface antigen


  Query Match             53.8%;  Score 49;  DB 2;  Length 527;
  Best Local Similarity   62.5%;  Pred. No. 15;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWW 8
           :|: ||||
Db     106 VLRNWPWW 113


RESULT   7
B70741
probable moeY protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: B70741
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: B70741
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-715 <COL>
A;Cross-references: GB:Z75555; GB:AL123456; NID:g3261608; PIDN:CAA99988.1; PID:e25035
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: moeY


  Query Match             53.8%;  Score 49;  DB 2;  Length 715;
  Best Local Similarity   60.0%;  Pred. No. 20;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 KKWPWWPWRR 12
           |:| ::||||
Db      64 KRWAYYPWRR 73


RESULT   8
A59287
myosin heavy chain - fluke (Schistosoma mansoni) (strain Brazilian LE)
C;Species: Schistosoma mansoni
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59287
R;Weston, D.S.; Schmitz, J.; Kemp, M.; Kunz, W.
Mol. Biochem. Parasitol. 58, 161-164, 1993
A;Title: Cloning and sequence characterization of a complete myosin heavy chain cDNA
A;Reference number: A59287; MUID:93211444
A;Accession: A59287
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1940 <WES>
A;Cross-references: GB:L01634; PIDN:AAA29905.1
A;Experimental source: strain Brazilian LE
C;Genetics:
A;Gene: MYH
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;82-752/Domain: myosin motor domain homology <MMO>


  Query Match             53.8%;  Score 49;  DB 2;  Length 1940;
  Best Local Similarity   62.5%;  Pred. No. 53;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ILKKWPWW 8
           :|: ||||
Db     809 VLRNWPWW 816


RESULT   9
T29295
hypothetical protein C50F7.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29295
R;Johnson, D.; Stellyes, L.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C50F7.
A;Reference number: Z20601
A;Accession: T29295
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-111 <JOH>
A;Cross-references: EMBL:U41557; PIDN:AAA83303.1; CESP:C50F7.8
C;Genetics:
```

```
A;Gene: CESP:C50F7.8

  Query Match             52.7%;  Score 48;  DB 2;  Length 111;
  Best Local Similarity   75.0%;  Pred. No. 4.6;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 WPWWPWRR 12
           |||||  |
Db      15 WPWWPGGR 22


RESULT  10
T36208
hypothetical protein SCE36.09 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36208
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z21601
A;Accession: T36208
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-114 <OLI>
A;Cross-references: EMBL:AL049763; PIDN:CAB42078.1; GSPDB:GN00070; SCOEDB:SCE36.09
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE36.09

  Query Match             52.2%;  Score 47.5;  DB 2;  Length 114;
  Best Local Similarity   63.6%;  Pred. No. 5.5;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       3 KKW-PWWPWRR 12
           ::| | |||||
Db     102 RRWRPRWPWRR 112


RESULT  11
S23449
NADH oxidase (H2O2-forming) (EC 1.6.-.-) - Thermus aquaticus
C;Species: Thermus aquaticus
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 23-Mar-1993
C;Accession: S23449; S24556
R;Park, H.J.; Kreutzer, R.; Reiser, C.O.A.; Sprinzl, M.
Eur. J. Biochem. 205, 875-879, 1992
A;Title: Molecular cloning and nucleotide sequence of the gene encoding a H(2)O(2)-formi
A;Reference number: S23449; MUID:92249331
A;Accession: S23449
A;Molecule type: DNA
A;Residues: 1-248 <PAR>
A;Cross-references: EMBL:X60110
A;Accession: S24556
A;Molecule type: protein
A;Residues: 1-32 <PAR1>
C;Genetics:
A;Gene: nox
C;Keywords: NAD; oxidoreductase
F;1-248/Product: NADH oxidase (H2O2-forming) #status experimental <MAT>

  Query Match             51.6%;  Score 47;  DB 2;  Length 248;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 PWWPW 10
           |||||
Db     179 PWWPW 183


RESULT  12
G70715
hypothetical protein Rv0945 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70715
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: G70715
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-253 <COL>
A;Cross-references: GB:Z79700; GB:AL123456; NID:g3261628; PIDN:CAB02005.1; PID:g15242
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0945
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F;8-190/Domain: short-chain alcohol dehydrogenase homology <SADH>

  Query Match             51.6%;  Score 47;  DB 2;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 PWWPW 10
           |||||
Db     230 PWWPW 234


RESULT  13
S77354
histidinol-phosphate aminotransferase hisC-1 - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein sll1713
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S77354
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S77354
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-352 <KAN>
A;Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAA17457.1; PID:g165
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: hisC-1
C;Superfamily: probable histidinol-phosphate transaminase

  Query Match             51.1%;  Score 46.5;  DB 2;  Length 352;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    5;  Gaps    1;

Qy       2 LKKWPW-----WPWRR 12
           ||  ||     ||||:
Db     106 LKTIPWQVDQPWPWRK 121


RESULT  14
S37664
peplomeric polyprotein precursor - avian infectious bronchitis virus (strain D1466) (
N;Contains: E2 glycoprotein subunit S2
C;Species: avian infectious bronchitis virus, IBV
A;Variety: strain D1466
```

```
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
C;Accession: S37664
R;Kusters, J.G.; Jager, E.J.; Niesters, H.G.M.; van der Zeijst, B.A.M.
Vaccine 8, 605-608, 1990
A;Title: Sequence evidence for RNA recombination in field isolates of avian coronavirus
A;Reference number: S37663; MUID:91205880
A;Accession: S37664
A;Molecule type: genomic RNA
A;Residues: 1-621 <KUS>
A;Cross-references: EMBL:X58001; NID:g58986; PIDN:CAA41065.1; PID:g58987
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; peplomer protein; spike protein
F;1-5/Product: E2 glycoprotein subunit S1 (fragment) #status predicted <GS1>
F;6-621/Product: E2 glycoprotein subunit S2 #status predicted <GS2>
F;10,47,59,137,144,415,447,482,506,519,542/Binding site: carbohydrate (Asn) (covalent) #


  Query Match             51.1%;  Score 46.5;  DB 2;  Length 621;
  Best Local Similarity   61.5%;  Pred. No. 38;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
           |||   |||| : |
Db     553 ILKTYIKWPWYVW 565


RESULT  15
S37663
peplomeric polyprotein precursor - avian infectious bronchitis virus (strain D207) (frag
N;Contains: E2 glycoprotein subunit S2
C;Species: avian infectious bronchitis virus, IBV
A;Variety: strain D207
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
C;Accession: S37663
R;Kusters, J.G.; Jager, E.J.; Niesters, H.G.M.; van der Zeijst, B.A.M.
Vaccine 8, 605-608, 1990
A;Title: Sequence evidence for RNA recombination in field isolates of avian coronavirus
A;Reference number: S37663; MUID:91205880
A;Accession: S37663
A;Molecule type: genomic RNA
A;Residues: 1-630 <KUS>
A;Cross-references: EMBL:X58003; NID:g58988; PIDN:CAA41067.1; PID:g58989
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; peplomer protein; spike protein
F;1-5/Product: E2 glycoprotein subunit S1 (fragment) #status predicted <GS1>
F;6-621/Product: E2 glycoprotein subunit S2 #status predicted <GS2>
F;10,47,59,137,144,415,447,482,506,519,542/Binding site: carbohydrate (Asn) (covalent) #


  Query Match             51.1%;  Score 46.5;  DB 2;  Length 630;
  Best Local Similarity   61.5%;  Pred. No. 39;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    1;

Qy       1 ILK---KWPWWPW 10
           |||   |||| : |
Db     553 ILKTYIKWPWYVW 565


Search completed: January  4, 2002, 08:41:32
Job time: 170 sec
```

```
OM of: US-09-444-281-36 to: EST:*   out_format : pfs

Date: Jan 4, 2002 10:25 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084143_16183/app_query.fasta_1.210
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09444281_@CGN1_1_3354 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-36
Query length: 12
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 2868.270000


score_list:
Sequence              Strd Orig   ZScore   EScore Len   ! Documentation  ..
gb_est1:AU198144       -    69.00  142.59  88.74     446  ! AU198144 AU198144 Rice green sh
gb_est1:AU198162       -    69.00  142.56  89.13     448  ! AU198162 AU198162 Rice green sh
gb_est1:AU082117       -    69.00  140.66  113.72    578  ! AU082117 AU082117 Rice panicle
gb_est1:AU089922       -    68.00  140.66  113.72    446  ! AU089922 AU089922 Hordeum vulga
gb_est1:AU089934       -    68.00  140.19  120.78    475  ! AU089934 AU089934 Hordeum vulga
gb_est1:BE024584       +    67.00  140.97  109.24    330  ! BE024584 894003H02.y1 C. reinha
gb_est1:AU198258       -    67.00  140.49  116.19    352  ! AU198258 AU198258 Rice green sh
gb_est1:AV641634       +    67.00  140.10  122.19    371  ! AV641634 AV641634 Chlamydomonas
gb_est1:BE129188       +    67.00  139.80  126.91    386  ! BE129188 894021E12.y1 C. reinha
gb_est2:BF584442       +    67.00  135.35  224.57    701  ! BF584442 602098336F1 NCI_CGAP_C
gb_est2:BG964576       +    67.00  134.05  265.17    834  ! BG964576 602832255F1 NCI_CGAP_C
gb_gss:AQ943724        -    66.00  135.73  213.87    514  ! AQ943724 Sheared DNA-36H6.TF Sh
gb_gss:AZ218599        -    66.00  134.93  236.90    572  ! AZ218599 Sheared DNA-82B1.TR Sh
gb_est2:C52835         +    65.00  140.55  115.35    208  ! C52835 C52835 Yuji Kohara unpub
gb_est2:C62320         -    65.00  138.02  159.56    292  ! C62320 C62320 Yuji Kohara unpub
gb_est1:BE237369       +    65.00  135.39  223.33    415  ! BE237369 146629 MARC 4BOV Bos t
gb_est1:AU090148       +    65.00  134.96  236.19    440  ! AU090148 AU090148 Hordeum vulga
gb_est1:AU090136       +    65.00  134.66  245.42    458  ! AU090136 AU090136 Hordeum vulga
gb_est2:BI174860       +    65.00  132.64  317.76    600  ! BI174860 OSTF061C5_1 AD-wrmcDNA
gb_est1:BE422636       -    64.00  133.51  284.20    412  ! BE422636 WHE0057_F03_L05ZS Whea
gb_est1:BE494459       +    64.00  133.42  287.50    417  ! BE494459 WHE1254_A02_B04ZS Seca
gb_est1:BE453064       -    64.00  133.13  298.70    434  ! BE453064 894068B05.x1 C. reinha
gb_est1:AW762881       -    64.00  132.06  342.67    501  ! AW762881 ur59b09.y1 NCI_CGAP_Ma
gb_gss:AQ448482        -    64.00  131.71  358.35    525  ! AQ448482 mgxb0020D20f CUGI Rice
gb_est1:BE123173       -    63.50  135.24  227.65    287  ! BE123173 945024C07.Y2 945 - Mix
gb_est1:BE123317       -    63.50  131.97  346.32    445  ! BE123317 945041B02.Y1 945 - Mix
gb_est1:AW424548       -    63.50  131.84  352.27    453  ! AW424548 707017F11.y1 707 - Mix
gb_est2:BG901201       -    63.50  131.76  355.99    458  ! BG901201 950001B11.x2 950 - Mat
gb_est1:BE643645       -    63.50  131.42  371.59    479  ! BE643645 945041B06.Y2 945 - Mix
gb_est1:AW562908       -    63.50  131.06  389.38    503  ! AW562908 660069E08.x1 660 - Mix
gb_est1:BE123319       -    63.50  130.88  398.26    515  ! BE123319 945041B06.Y1 945 - Mix
gb_est1:AW562591       +    63.50  130.77  404.17    523  ! AW562591 660066G07.y1 660 - Mix
gb_est1:BE643603       -    63.50  130.70  407.87    528  ! BE643603 945024C07.Y3 945 - Mix
gb_est1:BE224959       +    63.50  130.61  412.30    534  ! BE224959 945041B02.x2 945 - Mix
gb_est1:BE643643       -    63.50  130.57  414.52    537  ! BE643643 945041B02.Y2 945 - Mix
gb_est1:BE129769       +    63.50  130.38  424.85    551  ! BE129769 945024C07.X2 945 - Mix
gb_est1:BE643551       +    63.50  130.27  430.75    559  ! BE643551 945041B02.X1 945 - Mix
gb_est1:BE643554       +    63.50  130.26  431.48    560  ! BE643554 945041B06.X1 945 - Mix
gb_est1:AW355984       +    63.50  130.14  438.11    569  ! AW355984 707017F11.x1 707 - Mix
gb_est1:BE056153       -    63.50  130.14  438.11    569  ! BE056153 945024C07.Y1 945 - Mix
gb_est1:AW566321       +    63.50  129.98  446.95    581  ! AW566321 660069E08.y1 660 - Mix
gb_est2:BE977776       +    63.00  135.16  230.15    255  ! BE977776 bs68h10.y1 Drosophi
gb_est2:BG051614       -    63.00  134.38  254.26    283  ! BG051614 FM1_58_H09.g1_A003
gb_est2:BG305131       -    63.00  131.01  392.03    445  ! BG305131 fl95a09.y1 Zebrafis
gb_est2:BG711576       +    63.00  130.76  404.66    460  ! BG711576 pg11n.pk008.f16 Nor

seq_name: gb_est1:AU198144

seq_documentation_block:
LOCUS       AU198144      446 bp    mRNA            EST       12-JUL-2001
DEFINITION  AU198144 Rice green shoot Oryza sativa cDNA clone S15951, mRNA
            sequence.
ACCESSION   AU198144
VERSION     AU198144.1  GI:14714211
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT = 'RGP'.
            S15951_97A.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S15951"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT       82 a    153 c    161 g     50 t
ORIGIN


alignment_scores:
            Quality:   69.00             Length:     10
              Ratio:  7.667               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AU198144/rev   ..

 Align seg 1/1  to reverse of: AU198144  from: 1  to: 446

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         ||||||||||||||||||||   |||:::
     348 CGCTGGCCTTGGTGGCCCTGGACGCGGCGG 319

seq_name: gb_est1:AU198162

seq_documentation_block:
LOCUS       AU198162      448 bp    mRNA            EST       12-JUL-2001
DEFINITION  AU198162 Rice green shoot Oryza sativa cDNA clone S16019, mRNA
            sequence.
ACCESSION   AU198162
VERSION     AU198162.1  GI:14714231
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 448)
```

```
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT = 'RGP'.
            S16019_97A.
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S16019"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT       85 a    146 c    160 g     57 t
ORIGIN


alignment_scores:
            Quality:   69.00             Length:     10
              Ratio:  7.667               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AU198162/rev   ..

 Align seg 1/1  to reverse of: AU198162  from: 1  to: 448

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         |||||||||||||||||||   |||:::
     311 CGCTGGCCTTGGTGGCCCTGGACGCGGCGG 282

seq_name: gb_est1:AU082117

seq_documentation_block:
LOCUS       AU082117      578 bp    mRNA            EST       04-FEB-2000
DEFINITION  AU082117 Rice panicle at ripening stage Oryza sativa cDNA clone
            E11611, mRNA sequence.
ACCESSION   AU082117
VERSION     AU082117.1  GI:6727452
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle at ripening stage (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E11611"
                     /clone_lib="Rice panicle at ripening stage"
                     /dev_stage="ripening stage"
                     /note="Organ: panicle; Rice cDNA from panicle at ripening
                     stage"
BASE COUNT      129 a    180 c    179 g     90 t
ORIGIN


alignment_scores:
            Quality:   69.00             Length:     10
              Ratio:  7.667               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AU082117/rev   ..

 Align seg 1/1  to reverse of: AU082117  from: 1  to: 578

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         |||||||||||||||||||   |||:::
     345 CGCTGGCCTTGGTGGCCCTGGACGCGGCGG 316

seq_name: gb_est1:AU089922

seq_documentation_block:
LOCUS       AU089922      446 bp    mRNA            EST       19-APR-2000
DEFINITION  AU089922 Hordeum vulgare subsp. vulgare Upper three leaves at
            heading stage Hordeum vulgare subsp. vulgare cDNA clone
            haruna_lib1_121, mRNA sequence.
ACCESSION   AU089922
VERSION     AU089922.1  GI:7613350
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Sato,K., Takahashi,H. and Takeda,K.
  TITLE     Hordeum vulgare subsp. vulgare cDNA clone
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Kazuhiro Sato
            Research Institute for Bioresources
            Okayama University, Barley Germplasm Center
            Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
            Email: kazsato@rib.okayama-u.ac.jp,
            URL:http://www.rib.okayama-u.ac.jp/barley/.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="haruna_lib1_121"
                     /clone_lib="Hordeum vulgare subsp. vulgare Upper three
                     leaves at heading stage"
                     /tissue_type="Upper three leaves at heading stage"
BASE COUNT       89 a    130 c    149 g     76 t      2 others
ORIGIN


alignment_scores:
            Quality:   68.00             Length:      8
              Ratio:  8.500               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-36 x AU089922/rev   ..

 Align seg 1/1  to reverse of: AU089922  from: 1  to: 446

       4 TrpProTrpTrpProTrpArgArg 11
         ||||||||||||||||||||||||
     188 TGGCCGTGGTGGCCGTGGCGGCGA 165

seq_name: gb_est1:AU089934
```

```
seq_documentation_block:
LOCUS       AU089934      475 bp    mRNA            EST       19-APR-2000
DEFINITION  AU089934 Hordeum vulgare subsp. vulgare Upper three leaves at
            heading stage Hordeum vulgare subsp. vulgare cDNA clone
            haruna_lib1_134, mRNA sequence.
ACCESSION   AU089934
VERSION     AU089934.1  GI:7613362
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Sato,K., Takahashi,H. and Takeda,K.
  TITLE     Hordeum vulgare subsp. vulgare cDNA clone
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Kazuhiro Sato
            Research Institute for Bioresources
            Okayama University, Barley Germplasm Center
            Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
            Email: kazsato@rib.okayama-u.ac.jp,
            URL:http://www.rib.okayama-u.ac.jp/barley/.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="haruna_lib1_134"
                     /clone_lib="Hordeum vulgare subsp. vulgare Upper three
                     leaves at heading stage"
                     /tissue_type="Upper three leaves at heading stage"
BASE COUNT      100 a    137 c    155 g     83 t
ORIGIN


alignment_scores:
            Quality:   68.00             Length:      8
              Ratio:  8.500               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-36 x AU089934/rev   ..

 Align seg 1/1  to reverse of: AU089934  from: 1  to: 475

       4 TrpProTrpTrpProTrpArgArg 11
         ||||||||||||||||||||||||
     194 TGGCCGTGGTGGCCGTGGCGGCGA 171

seq_name: gb_est1:BE024584

seq_documentation_block:
LOCUS       BE024584      330 bp    mRNA            EST       06-JUN-2000
DEFINITION  894003H02.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE024584
VERSION     BE024584.1  GI:8287025
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      40 a    103 c    142 g     45 t
ORIGIN


alignment_scores:
            Quality:   67.00             Length:      9
              Ratio:  8.375               Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x BE024584   ..

 Align seg 1/1  to: BE024584  from: 1  to: 330

       3 ArgTrpProTrpTrpProTrpArgArg 11
         ||||||||||||||||||||||   |||
      72 CGGTGGCCGTGGTGGCCGTGGGCGCGG 98

seq_name: gb_est1:AU198258

seq_documentation_block:
LOCUS       AU198258      352 bp    mRNA            EST       12-JUL-2001
DEFINITION  AU198258 Rice green shoot Oryza sativa cDNA clone S16389, mRNA
            sequence.
ACCESSION   AU198258
VERSION     AU198258.1  GI:14714335
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT = 'RGP'.
            S16389_96Z.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
```

```
                     /db_xref="taxon:4530"
                     /clone="S16389"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT        95 a     95 c    104 g     51 t      7 others
ORIGIN


alignment_scores:
            Quality:   67.00          Length:      9
              Ratio:  8.375             Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AU198258/rev    ..

 Align seg 1/1  to reverse of: AU198258  from: 1  to: 352

       3 ArgTrpProTrpTrpProTrpArgArg 11
         ||||||||||||||||||||   |||
     167 CGCTGGCCTTGGTGGCCCTGGACGCGG 141

seq_name: gb_est1:AV641634

seq_documentation_block:
LOCUS       AV641634      371 bp    mRNA            EST       15-DEC-2000
DEFINITION  AV641634 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HCL037h03_r 5', mRNA sequence.
ACCESSION   AV641634
VERSION     AV641634.1  GI:10784962
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HCL037h03_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT        73 a    106 c    141 g     51 t
ORIGIN


alignment_scores:
            Quality:   67.00          Length:      9
              Ratio:  8.375             Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AV641634    ..

 Align seg 1/1  to: AV641634  from: 1  to: 371

       3 ArgTrpProTrpTrpProTrpArgArg 11
         ||||||||||||||||||||   |||
     244 CGGTGGCCGTGGTGGCCGTGGGCGCGG 270

seq_name: gb_est1:BE129188

seq_documentation_block:
LOCUS       BE129188      386 bp    mRNA            EST       21-JUN-2000
DEFINITION  894021E12.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE129188
VERSION     BE129188.1  GI:8576551
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT        79 a    109 c    142 g     56 t
ORIGIN


alignment_scores:
            Quality:   67.00          Length:      9
              Ratio:  8.375             Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x BE129188    ..

 Align seg 1/1  to: BE129188  from: 1  to: 386

       3 ArgTrpProTrpTrpProTrpArgArg 11
         ||||||||||||||||||||   |||
     249 CGGTGGCCGTGGTGGCCGTGGGCGCGG 275

seq_name: gb_est2:BF584442

seq_documentation_block:
LOCUS       BF584442      701 bp    mRNA            EST       12-DEC-2000
DEFINITION  602098336F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4218273 5',
```

```
            mRNA sequence.
ACCESSION   BF584442
VERSION     BF584442.1  GI:11658160
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9798  row: g  column: 10
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4218273"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      149 a    198 c    186 g    168 t
ORIGIN


alignment_scores:
            Quality:   67.00             Length:     10
              Ratio:  7.444               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x BF584442  ..

 Align seg 1/1  to: BF584442  from: 1  to: 701

       1 IleLeuArgTrpProTrpTrpProTrpArg 10
         :::|||   ||||||||||||||||||||
     188 CTGTTATTGTGGCCTTGGTGGCCATGGAGA 217

seq_name: gb_est2:BG964576

seq_documentation_block:
LOCUS       BG964576      834 bp    mRNA            EST       12-JUN-2001
DEFINITION  602832255F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986828 5',
            mRNA sequence.
ACCESSION   BG964576
VERSION     BG964576.1  GI:14352213
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10996  row: n  column: 13
            High quality sequence stop: 811.
FEATURES             Location/Qualifiers
     source          1. .834
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4986828"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      185 a    202 c    226 g    221 t
ORIGIN


alignment_scores:
            Quality:   67.00             Length:     10
              Ratio:  7.444               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x BG964576  ..

 Align seg 1/1  to: BG964576  from: 1  to: 834

       1 IleLeuArgTrpProTrpTrpProTrpArg 10
         :::|||   ||||||||||||||||||||
     195 CTGTTATTGTGGCCTTGGTGGCCATGGAGA 224

seq_name: gb_gss:AQ943724

seq_documentation_block:
LOCUS       AQ943724      514 bp    DNA             GSS       27-JAN-2000
DEFINITION  Sheared DNA-36H6.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-36H6, DNA sequence.
ACCESSION   AQ943724
VERSION     AQ943724.1  GI:6766989
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-36H6.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
```

```
    source          1. .514
                    /organism="Trypanosoma brucei"
                    /strain="TREU927/4 GUTat 10.1"
                    /db_xref="taxon:5691"
                    /clone="Sheared DNA-36H6"
                    /clone_lib="Sheared DNA"
                    /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                    Institute for Genomic Research (TIGR), Rockville, MD.
                    Genomic DNA isolated from a cloned population of
                    Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                    sheared to give a tight size distribution (approx 2 kb).
                    The v + i method used for the library construction is
                    described in detail in Smith, H.O. and Venter, J.C.
                    (Making small insert libraries for whole genome shotgun
                    sequencing projects.  In Genome Sequencing: A Practical
                    Approach, eds. M. Vaudin and B. Barell, Oxford University
                    Press, 1999)."
BASE COUNT      124 a    185 c    131 g     74 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     10
              Ratio:  7.333               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 90.000

alignment_block:
 US-09-444-281-36 x AQ943724/rev   ..

 Align seg 1/1  to reverse of: AQ943724  from: 1  to: 514

       2 LeuArgTrpProTrpTrpProTrpArgArg 11
         |||||||||||||||||   |||||||||
     450 CTTCGATGGCCTTGGTGGTGGTGGCGCCGG 421

seq_name: gb_gss:AZ218599

seq_documentation_block:
LOCUS        AZ218599      572 bp    DNA             GSS       09-JUN-2000
DEFINITION   Sheared DNA-82B1.TR Sheared DNA Trypanosoma brucei genomic clone
             Sheared DNA-82B1, DNA sequence.
ACCESSION    AZ218599
VERSION      AZ218599.1  GI:8436399
KEYWORDS     GSS.
SOURCE       Trypanosoma brucei.
  ORGANISM   Trypanosoma brucei
             Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
             Trypanosoma.
REFERENCE    1  (bases 1 to 572)
  AUTHORS    El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
             Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
             , Fraser,C. and Adams,M.
  TITLE      Determination of clone end sequences from Trypanosoma brucei GUTat
             10.1 sheared DNA library
  JOURNAL    Unpublished (1999)
COMMENT      Other_GSSs: Sheared DNA-82B1.TF
             Contact: Najib M. El-Sayed
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: nelsayed@tigr.org
             Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
             DNA library constructed at TIGR. Clones will be available for
             distribution through Research Genetics, Alabama, USA. Sheared DNA
             end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
             Seq primer: M13-Reverse
             Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .572
                      /organism="Trypanosoma brucei"
                      /strain="TREU927/4 GUTat 10.1"
                      /db_xref="taxon:5691"
                      /clone="Sheared DNA-82B1"
                      /clone_lib="Sheared DNA"
                      /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                      Institute for Genomic Research (TIGR), Rockville, MD.
                      Genomic DNA isolated from a cloned population of
                      Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                      sheared to give a tight size distribution (approx 2 kb).
                      The v + i method used for the library construction is
                      described in detail in Smith, H.O. and Venter, J.C.
                      (Making small insert libraries for whole genome shotgun
                      sequencing projects.  In Genome Sequencing: A Practical
                      Approach, eds. M. Vaudin and B. Barell, Oxford University
                      Press, 1999)."
BASE COUNT      136 a    217 c    144 g     75 t
ORIGIN


alignment_scores:
            Quality:   66.00             Length:     10
              Ratio:  7.333               Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 90.000

alignment_block:
 US-09-444-281-36 x AZ218599/rev   ..

 Align seg 1/1  to reverse of: AZ218599  from: 1  to: 572

       2 LeuArgTrpProTrpTrpProTrpArgArg 11
         |||||||||||||||||   |||||||||
     211 CTTCGATGGCCTTGGTGGTGGTGGCGCCGG 182

seq_name: gb_est2:C52835

seq_documentation_block:
LOCUS        C52835         208 bp    mRNA            EST       11-SEP-1997
DEFINITION   C52835 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
             clone yk285b1 3', mRNA sequence.
ACCESSION    C52835
VERSION      C52835.1  GI:2390592
KEYWORDS     EST.
SOURCE       Caenorhabditis elegans.
  ORGANISM   Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
             ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE    1  (bases 1 to 208)
  AUTHORS    Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
             ,M., Miyata,A. and Nishigaki,A.
  TITLE      Expression map of the C.elegans genome
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Yuji Kohara
             Genome Biology Lab.
             National Institute of Genetics
             Yata 1111, Mishima, Shizuoka 411, Japan
             Tel: 81-559-81-6854
             Fax: 81-559-81-6855
             Email: ykohara@lab.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .208
                      /organism="Caenorhabditis elegans"
                      /strain="CB1489 him-8(e1489)"
                      /db_xref="taxon:6239"
                      /clone="yk285b1"
                      /clone_lib="Yuji Kohara unpublished cDNA"
                      /sex="hermaphrodite, male"
                      /tissue_type="whole animal"
                      /dev_stage="varied"
BASE COUNT       25 a     50 c     77 g     52 t      4 others
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     10
```

```
            Ratio:  7.222                Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-36 x C52835   ..

 Align seg 1/1  to: C52835  from: 1  to: 208

       1 IleLeuArgTrpProTrpTrpProTrpArg 10
         ::::::   ||||||||||||||||||||
      47 GTGATGTGGTGGCCCTGGTGGCCCTGGCGG 76

seq_name: gb_est2:C62320

seq_documentation_block:
LOCUS       C62320          292 bp    mRNA            EST       22-SEP-1997
DEFINITION  C62320 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk285b1 5', mRNA sequence.
ACCESSION   C62320
VERSION     C62320.1  GI:2421025
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk285b1"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT       94 a     89 c     60 g     49 t
ORIGIN


alignment_scores:
            Quality:   65.00             Length:     10
              Ratio:  7.222                Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
 US-09-444-281-36 x C62320/rev   ..

 Align seg 1/1  to reverse of: C62320  from: 1  to: 292

       1 IleLeuArgTrpProTrpTrpProTrpArg 10
         ::::::   ||||||||||||||||||||
     163 GTGATGTGGTGGCCCTGGTGGCCCTGGCGG 134
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:47:24 ; Search time 50.17 Seconds
(without alignments)
34.986 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRWPWWPWRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	57	66.3	723	12	Q9DUC4	Q9duc4 tt virus. o
2	53	61.6	746	12	Q9JH31	Q9jh31 tt virus. o
3	53	61.6	1173	12	Q990M4	Q990m4 human coron
4	53	61.6	1173	12	Q990M3	Q990m3 human coron
5	53	61.6	1173	12	Q990M2	Q990m2 human coron
6	53	61.6	1173	12	Q990M1	Q990m1 human coron
7	53	61.6	1383	12	Q84712	Q84712 porcine epi
8	52	60.5	1245	3	Q9Y7V5	Q9y7v5 trichoderma
9	51	59.3	299	4	Q9Y4N1	Q9y4n1 homo sapien
10	51	59.3	504	2	P96143	P96143 thermoactin
11	50	58.1	111	5	Q18753	Q18753 caenorhabdi
12	50	58.1	141	11	Q9CZA1	Q9cza1 mus musculu
13	50	58.1	327	10	Q9AUN3	Q9aun3 oryza sativ
14	50	58.1	735	12	Q9DUC9	Q9duc9 tt virus. o
15	49	57.0	49	12	Q9DT80	Q9dt80 tt virus. o
16	49	57.0	226	4	Q9BSG8	Q9bsg8 homo sapien
17	49	57.0	467	5	Q19573	Q19573 caenorhabdi
18	49	57.0	748	12	Q9DT81	Q9dt81 tt virus. o
19	48.5	56.4	114	2	Q9X8C2	Q9x8c2 streptomyce
20	48	55.8	540	2	O07504	O07504 bacillus me
21	47	54.7	165	10	Q9SNN3	Q9snn3 oryza sativ
22	47	54.7	276	2	Q9HXC9	Q9hxc9 pseudomonas
23	47	54.7	1411	10	Q9LYG0	Q9lyg0 arabidopsis
24	46	53.5	154	2	Q9R6J3	Q9r6j3 agrobacteri
25	46	53.5	728	3	Q9P3G0	Q9p3g0 neurospora
26	45	52.3	159	2	Q9KZT3	Q9kzt3 streptomyce
27	45	52.3	273	2	Q9PCR3	Q9pcr3 xylella fas
28	45	52.3	412	2	Q9I6F7	Q9i6f7 pseudomonas
29	45	52.3	423	2	O24742	O24742 bacteroides
30	45	52.3	443	10	Q9S751	Q9s751 oryza sativ
31	45	52.3	448	2	Q9WYR8	Q9wyr8 thermotoga
32	45	52.3	525	10	Q9ATU5	Q9atu5 lolium rigi
33	45	52.3	525	10	Q9ATU2	Q9atu2 lolium rigi
34	45	52.3	525	10	Q9ATU1	Q9atu1 lolium rigi
35	45	52.3	730	10	Q9FGZ6	Q9fgz6 arabidopsis
36	45	52.3	767	12	Q9QUD8	Q9qud8 tt virus. h
37	45	52.3	1100	11	Q9Z1J9	Q9z1j9 mus musculu
38	44.5	51.7	766	12	Q9IFV0	Q9ifv0 tt virus. p
39	44	51.2	143	4	Q9H9A4	Q9h9a4 homo sapien
40	44	51.2	145	2	O86437	O86437 pseudomonas
41	44	51.2	257	2	Q56924	Q56924 yersinia en
42	44	51.2	361	4	Q9P1W6	Q9p1w6 homo sapien
43	44	51.2	406	5	Q9W404	Q9w404 drosophila
44	44	51.2	429	5	Q9N8Y2	Q9n8y2 trypanosoma
45	44	51.2	458	4	Q9UGB1	Q9ugb1 homo sapien

ALIGNMENTS

RESULT 1
Q9DUC4
ID Q9DUC4 PRELIMINARY; PRT; 723 AA.
AC Q9DUC4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RA Okamoto H.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RX PubMed=11080484;
RA Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,
RA Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;
RT "Species-specific TT viruses in humans and nonhuman primates and their
RT phylogenetic relatedness.";
RL Virology 277:368-378(2000).
DR EMBL; AB041959; BAB19313.1; -.
DR InterPro; IPR001563; Serine_carbpept.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; UNKNOWN_1.
SQ SEQUENCE 723 AA; 85393 MW; 232D003098766344 CRC64;

Query Match 66.3%; Score 57; DB 12; Length 723;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       5 PWWPWRR 11
            |||||||
Db       2 PWWPWRR 8
```

RESULT 2
Q9JH31

```
ID   Q9JH31      PRELIMINARY;      PRT;   746 AA.
AC   Q9JH31;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ORF1.
OS   TT virus.
OC   Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX   NCBI_TaxID=68887;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TJN02;
RA   Okamoto H.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TJN02;
RA   Ukita M., Okamoto H., Nishizawa T., Tawara A., Takahashi M.,
RA   Iizuka H., Miyakawa Y., Mayumi M.;
RT   "The entire nucleotide sequences of two distinct TT virus (TTV)
RT   isolates (TJN01 and TJN02) remotely related to the original TTV
RT   isolates.";
RL   Arch. Virol. 0:0-0(2000).
DR   EMBL; AB028669; BAA94878.1; -.
SQ   SEQUENCE   746 AA;  88561 MW;  E0B22953AE764E3E CRC64;


  Query Match             61.6%;  Score 53;  DB 12;  Length 746;
  Best Local Similarity   54.5%;  Pred. No. 15;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPWRRK 12
          : | || |||:
Db      1 MAWGWWRWRRR 11


RESULT   3
Q990M4
ID   Q990M4      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344186; AAK32188.1; -.
SQ   SEQUENCE   1173 AA;  128669 MW;  ABC6E0A75E8BD8A4 CRC64;


  Query Match             61.6%;  Score 53;  DB 12;  Length 1173;
  Best Local Similarity   62.5%;  Pred. No. 23;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPW 9
          ::|||| |
Db   1112 IKWPWWVW 1119


RESULT   4
Q990M3
ID   Q990M3      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344187; AAK32189.1; -.
SQ   SEQUENCE   1173 AA;  128683 MW;  9E2368160082A81A CRC64;


  Query Match             61.6%;  Score 53;  DB 12;  Length 1173;
  Best Local Similarity   62.5%;  Pred. No. 23;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPW 9
          ::|||| |
Db   1112 IKWPWWVW 1119


RESULT   5
Q990M2
ID   Q990M2      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344188; AAK32190.1; -.
SQ   SEQUENCE   1173 AA;  128653 MW;  8B658FCBBD1842DA CRC64;


  Query Match             61.6%;  Score 53;  DB 12;  Length 1173;
  Best Local Similarity   62.5%;  Pred. No. 23;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPW 9
          ::|||| |
Db   1112 IKWPWWVW 1119


RESULT   6
Q990M1
ID   Q990M1      PRELIMINARY;      PRT;  1173 AA.
AC   Q990M1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE GLYCOPROTEIN.
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
```

```
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=229E;
RA   Bonavia A., Holmes K.V.;
RT   "Viral and cellular changes in a human cell line persistently infected
RT   with human coronavirus HCoV-229E.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF344189; AAK32191.1; -.
SQ   SEQUENCE   1173 AA;  128760 MW;  B73A165A6270152A CRC64;


  Query Match             61.6%;  Score 53;  DB 12;  Length 1173;
  Best Local Similarity   62.5%;  Pred. No. 23;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPW 9
          ::|||| |
Db   1112 IKWPWWVW 1119


RESULT   7
Q84712
ID   Q84712      PRELIMINARY;      PRT;  1383 AA.
AC   Q84712;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPIKE PROTEIN.
OS   porcine epidemic diarrhea virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=28295;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1/87;
RX   MEDLINE=94231173; PubMed=8176382;
RA   Duarte M., Laude H.;
RT   "Sequence of the spike protein of the porcine epidemic diarrhoea
RT   virus.";
RL   J. Gen. Virol. 75:1195-1200(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1/87;
RX   MEDLINE=93389433; PubMed=8397280;
RA   Bridgen A., Duarte M., Tobler K., Laude H., Ackermann M.;
RT   "Sequence determination of the nucleocapsid protein gene of the
RT   porcine epidemic diarrhoea virus confirms that this virus is a
RT   coronavirus related to human coronavirus 229E and porcine
RT   transmissible gastroenteritis virus.";
RL   J. Gen. Virol. 74:1795-1804(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BR1/87;
RX   MEDLINE=94120721; PubMed=8291230;
RA   Duarte M., Tobler K., Bridgen A., Rasschaert D., Ackermann M.,
RA   Laude H.;
RT   "Sequence analysis of the porcine epidemic diarrhea virus genome
RT   between the nucleocapsid and spike protein genes reveals a polymorphic
RT   ORF.";
RL   Virology 198:466-476(1994).
DR   EMBL; Z25483; CAA80971.1; -.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
FT   CONFLICT    422    422       Y -> N (IN REF. 1).
SQ   SEQUENCE   1383 AA;  151404 MW;  741C84D5DD3BDC4D CRC64;


  Query Match             61.6%;  Score 53;  DB 12;  Length 1383;
  Best Local Similarity   62.5%;  Pred. No. 27;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPW 9
          ::|||| |
Db   1321 IKWPWWVW 1328


RESULT   8
Q9Y7V5
ID   Q9Y7V5      PRELIMINARY;      PRT;  1245 AA.
AC   Q9Y7V5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CONIDIOSPORE SURFACE PROTEIN.
GN   CMP1.
OS   Trichoderma harzianum.
OC   Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Trichoderma.
OX   NCBI_TaxID=5544;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 32173;
RA   Puyesky M., Benhamou N., Ponce Noyola P., Bauw G., Ziv T.,
RA   van Montagu M., Herrera Estrella A., Horwitz B.A.;
RT   "Developmental regulation of a gene encoding a multidomain
RT   conidiospore surface protein of Trichoderma, cmp1.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ133651; CAB40845.1; -.
SQ   SEQUENCE   1245 AA;  135824 MW;  3249C749AFA0CDF8 CRC64;


  Query Match             60.5%;  Score 52;  DB 3;  Length 1245;
  Best Local Similarity   60.0%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 RWPWWPWRRK 12
          || || | |:
Db   1185 RWQWWSWPRR 1194


RESULT   9
Q9Y4N1
ID   Q9Y4N1      PRELIMINARY;   PRT;   299 AA.
AC   Q9Y4N1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 34.0 KDA PROTEIN (FRAGMENT).
GN   DKFZP434C192.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Ansorge W., Wirkner U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL096753; CAB46428.2; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   299 AA;  34032 MW;  6B8DB60E6A88239A CRC64;


  Query Match             59.3%;  Score 51;  DB 4;  Length 299;
  Best Local Similarity   85.7%;  Pred. No. 12;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 PWWPWRR 11
          ||||| |
Db     37 PWWPWSR 43
```

```
RESULT  10
P96143
ID   P96143      PRELIMINARY;      PRT;   504 AA.
AC   P96143;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PEPTIDE HYDROLASE.
GN   TLPE1.
OS   Thermoactinomyces vulgaris.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Thermoactinomyces.
OX   NCBI_TaxID=2026;
RN   [1]
RP   SEQUENCE OF 1-431 FROM N.A.
RC   STRAIN=94-2A;
RA   Hofemeister J.W.;
RL   Thesis (1995), Molecular Genetics,
RL   Institut fuer Pflanzengenetik und Kulturpflanzenforschung.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=94-2A;
RA   Hofemeister J.W.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z83214; CAB05671.1; -.
DR   HSSP; P00800; 1HYT.
DR   InterPro; IPR001570; Peptidase_M4.
DR   Pfam; PF01447; Peptidase_M4; 1.
KW   Hydrolase.
SQ   SEQUENCE   504 AA;  56653 MW;  5A7BCC05C5AD1315 CRC64;


  Query Match             59.3%;  Score 51;  DB 2;  Length 504;
  Best Local Similarity   60.0%;  Pred. No. 20;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWR 10
           ::|| ||| |
Db      71 LVRWTWWPMR 80


RESULT  11
Q18753
ID   Q18753      PRELIMINARY;      PRT;   111 AA.
AC   Q18753;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCINE-RICH.
GN   C50F7.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Johnson D., Stellyes L.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U41557; AAA83303.1; -.
SQ   SEQUENCE   111 AA;  10139 MW;  6E729A2E0F9762B9 CRC64;


  Query Match             58.1%;  Score 50;  DB 5;  Length 111;
  Best Local Similarity   54.5%;  Pred. No. 7;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRR 11
           :: |||||  |
Db      12 VMWWPWWPGGR 22


RESULT  12
Q9CZA1
ID   Q9CZA1      PRELIMINARY;      PRT;   141 AA.
AC   Q9CZA1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2810031J10RIK PROTEIN.
GN   2810031J10RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK012846; BAB28508.1; -.
DR   MGD; MGI:1919917; 2810031J10Rik.
DR   InterPro; IPR003309; SCAN.
DR   Pfam; PF02023; SCAN; 1.
DR   SMART; SM00431; LER; 1.
DR   PROSITE; PS50804; SCAN_BOX; 1.
SQ   SEQUENCE   141 AA;  15993 MW;  865C6B735BF8203D CRC64;


  Query Match             58.1%;  Score 50;  DB 11;  Length 141;
  Best Local Similarity   66.7%;  Pred. No. 8.7;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPW 9
           : | |||||
Db     105 VSRRPWWPW 113


RESULT  13
```

```
Q9AUN3
ID   Q9AUN3      PRELIMINARY;      PRT;   327 AA.
AC   Q9AUN3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Spiegel L.A., King L., Kirchoff K.A., de la Bastide M., Preston R.R.,
RA   Nascimento L.U., Vil M.D., Baker J.P., Miller B., Cunnius D.M.,
RA   Kuit K.H., Rodriguez S., Santos L., Zutavern T., Balija V.S.,
RA   Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
RA   McCombie W.R.;
RT   "Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
RT   Clone OSJNBa0058E19, Complete Sequence.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC083945; AAK13143.1; -.
SQ   SEQUENCE   327 AA;  36672 MW;  5CCA9080664BD0CA CRC64;


  Query Match             58.1%;  Score 50;  DB 10;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 WWPWRR 11
           ||||||
Db     119 WWPWRR 124


RESULT  14
Q9DUC9
ID   Q9DUC9      PRELIMINARY;      PRT;   735 AA.
AC   Q9DUC9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ORF1.
OS   TT virus.
OC   Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX   NCBI_TaxID=68887;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PT-TTV6;
RA   Okamoto H.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PT-TTV6;
RX   PubMed=11080484;
RA   Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,
RA   Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;
RT   "Species-specific TT viruses in humans and nonhuman primates and their
RT   phylogenetic relatedness.";
RL   Virology 277:368-378(2000).
DR   EMBL; AB041957; BAB19308.1; -.
SQ   SEQUENCE   735 AA;  86132 MW;  9ED818D6BE6FA5D3 CRC64;


  Query Match             58.1%;  Score 50;  DB 12;  Length 735;
  Best Local Similarity   46.7%;  Pred. No. 37;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    4;  Gaps    1;

Qy       2 LRWPW----WPWRRK 12
           : |||    | |||:
Db       1 MAWPWRRRRWRWRRR 15


RESULT  15
Q9DT80
ID   Q9DT80      PRELIMINARY;      PRT;   49 AA.
AC   Q9DT80;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ORF1 (FRAGMENT).
OS   TT virus.
OC   Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX   NCBI_TaxID=68887;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYM9;
RX   MEDLINE=20568739; PubMed=11118348;
RA   Okamoto H., Nishizawa T., Tawara A., Takahashi M., Kishimoto J.,
RA   Sai T., Sugai Y.;
RT   "TT virus mRNAs detected in the bone marrow cells from an infected
RT   individual.";
RL   Biochem. Biophys. Res. Commun. 279:700-707(2000).
DR   EMBL; AB050449; BAB19930.1; -.
FT   NON_TER      49     49
SQ   SEQUENCE   49 AA;  7225 MW;  1DA6F8F1AB69AA43 CRC64;


  Query Match             57.0%;  Score 49;  DB 12;  Length 49;
  Best Local Similarity   36.8%;  Pred. No. 4.6;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    8;  Gaps    1;

Qy       2 LRWPW--------WPWRRK 12
           : | |        |||||:
Db       1 MAWTWWWQRRRRRWPWRRR 19


Search completed: January  4, 2002, 08:47:26
Job time: 414 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:47:48 ; Search time 18.1 Seconds
(without alignments)
24.308 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRWPWWPWRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	70	81.4	144	1	INDC_BOVIN	P33046 bos taurus
2	53	61.6	1173	1	VGL2_CVH22	P15423 human coron
3	49	57.0	492	1	ADRO_BOVIN	P08165 bos taurus
4	47	54.7	253	1	Y945_MYCTU	P71564 mycobacteri
5	47	54.7	715	1	YD55_MYCTU	Q11025 mycobacteri
6	45.5	52.9	505	1	TRPE_PSESS	P21689 pseudomonas
7	45	52.3	196	1	YA05_SCHPO	Q09677 schizosacch
8	45	52.3	1108	1	CN3B_RAT	Q63085 rattus norv
9	44	51.2	361	1	FUT3_HUMAN	P21217 homo sapien
10	44	51.2	372	1	FUT3_PANTR	O19058 pan troglod
11	44	51.2	397	1	MML6_MYCTU	Q10773 mycobacteri
12	44	51.2	535	1	YDW6_SCHPO	O13912 schizosacch
13	44	51.2	967	1	MML4_MYCTU	O53735 mycobacteri
14	44	51.2	968	1	MML2_MYCTU	Q11171 mycobacteri
15	44	51.2	984	1	SX13_MOUSE	Q04891 mus musculu
16	44	51.2	1154	1	VGL2_IBVD2	P12722 avian infec
17	44	51.2	1162	1	VGL2_IBVB	P11223 avian infec
18	44	51.2	1162	1	VGL2_IBVK	P12650 avian infec
19	44	51.2	1162	1	VGL2_IBVM	P12651 avian infec
20	44	51.2	1163	1	VGL2_IBV6	P05135 avian infec
21	43.5	50.6	276	1	RCEL_RHOPA	O83005 rhodopseudo
22	43.5	50.6	2436	1	ABC2_HUMAN	Q9bzc7 homo sapien
23	43	50.0	51	1	LHB2_ECTHA	P80105 ectothiorho
24	43	50.0	711	1	MMLA_STRCO	Q53902 streptomyce
25	43	50.0	958	1	MML1_MYCTU	P95211 mycobacteri
26	43	50.0	1112	1	CN3B_HUMAN	Q13370 homo sapien
27	43	50.0	1225	1	VGL2_CVPR8	P27655 porcine res
28	43	50.0	1225	1	VGL2_CVPRM	P24413 porcine res
29	43	50.0	1235	1	VGL2_CVMJH	P11225 murine coro
30	43	50.0	1324	1	VGL2_CVMA5	P11224 murine coro
31	43	50.0	1353	1	VGL2_CVHOC	P36334 human coron
32	43	50.0	1363	1	VGL2_CVBF	P25190 bovine coro
33	43	50.0	1363	1	VGL2_CVBL9	P25191 bovine coro
34	43	50.0	1363	1	VGL2_CVBLY	P25192 bovine coro
35	43	50.0	1363	1	VGL2_CVBM	P15777 bovine coro
36	43	50.0	1363	1	VGL2_CVBQ	P25193 bovine coro
37	43	50.0	1363	1	VGL2_CVBV	P25194 bovine coro
38	43	50.0	1376	1	VGL2_CVM4	P22432 murine coro
39	43	50.0	1376	1	VGL2_CVMJC	Q02385 murine coro
40	43	50.0	1447	1	VGL2_CVPPR	Q02167 porcine tra
41	43	50.0	1447	1	VGL2_CVPPU	P07946 porcine tra
42	43	50.0	1447	1	VGL2_CVPRT	Q01977 porcine tra
43	43	50.0	1449	1	VGL2_CVPFS	P18450 porcine tra
44	43	50.0	1449	1	VGL2_CVPMI	P33470 porcine tra
45	43	50.0	1451	1	VGL2_CVCAI	P36300 canine ente

ALIGNMENTS

```
RESULT   1
INDC_BOVIN
ID   INDC_BOVIN     STANDARD;      PRT;   144 AA.
AC   P33046;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   INDOLICIDIN PRECURSOR.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=92392368; PubMed=1520337;
RA   del Sal G., Storici P., Schneider C., Romeo D., Zanetti M.;
RT   "cDNA cloning of the neutrophil bactericidal peptide indolicidin.";
RL   Biochem. Biophys. Res. Commun. 187:467-472(1992).
RN   [2]
RP   SEQUENCE OF 131-143.
RC   TISSUE=Neutrophils;
RX   MEDLINE=92165771; PubMed=1537821;
RA   Selsted M.E., Novotny M.J., Morris W.L., Tang Y.-Q., Smith W.,
RA   Cullor J.S.;
RT   "Indolicidin, a novel bactericidal tridecapeptide amide from
RT   neutrophils.";
RL   J. Biol. Chem. 267:4292-4295(1992).
CC   -!- FUNCTION: POTENT MICROBICIDAL ACTIVITY, ACTIVE AGAINST
CC       STAPHYLOCOCCUS AUREUS AND ESCHERICHIA COLI.
CC   -!- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.
CC   -!- PTM: ELASTASE MIGHT BE RESPONSIBLE FOR ITS MATURATION.
CC   -!- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67340; CAA47755.1; -.
DR   PIR; JC1222; JC1222.
DR   PIR; A42387; A42387.
DR   InterPro; IPR001894; Cathelicidin.
DR   Pfam; PF00666; Cathelicidins; 1.
DR   ProDom; PD001838; Cathelicidin; 1.
DR   PROSITE; PS00946; CATHELICIDINS_1; 1.
DR   PROSITE; PS00947; CATHELICIDINS_2; 1.
KW   Antibiotic; Amidation; Signal.
FT   SIGNAL       1     29       POTENTIAL.
FT   PROPEP      30    130
FT   PEPTIDE    131    143       INDOLICIDIN.
FT   MOD_RES     30     30       PYRROLIDONE CARBOXYLIC ACID (BY
```

```
FT                                SIMILARITY).
FT   DISULFID     85     96       BY SIMILARITY.
FT   DISULFID    107    124       BY SIMILARITY.
FT   MOD_RES     143    143       AMIDATION (G-144 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   144 AA;  16479 MW;  E3B1CBBE55C09911 CRC64;


  Query Match             81.4%;  Score 70;  DB 1;  Length 144;
  Best Local Similarity   88.9%;  Pred. No. 0.0039;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 RWPWWPWRR 11
          :|||||||||
Db    135 KWPWWPWRR 143


RESULT   2
VGL2_CVH22
ID   VGL2_CVH22     STANDARD;      PRT;  1173 AA.
AC   P15423;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
GN   S.
OS   Human coronavirus (strain 229E).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11137;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90264837; PubMed=2345367;
RA   Raabe T., Schelle-Prinz B., Siddell S.G.;
RT   "Nucleotide sequence of the gene encoding the spike glycoprotein of
RT   human coronavirus HCV 229E.";
RL   J. Gen. Virol. 71:1065-1073(1990).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC       AND IN SYNCYTIUM FORMATION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X16816; CAA34723.1; -.
DR   PIR; A34766; VGIHHC.
DR   InterPro; IPR002551; Corona_S1.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01600; Corona_S1; 1.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     15
FT   CHAIN        16   1173       E2 GLYCOPROTEIN.
FT   DOMAIN       16   1115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1116   1135       POTENTIAL.
FT   DOMAIN     1136   1173       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1136   1157       CYS-RICH.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     62     62       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    171    171       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    440    440       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    464    464       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    518    518       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    538    538       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    581    581       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    587    587       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    671    671       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    930    930       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1015   1015       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1020   1020       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1037   1037       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1066   1066       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1076   1076       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1082   1082       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1096   1096       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1173 AA;  128639 MW;  B9CA9A41A796B3BD CRC64;


  Query Match             61.6%;  Score 53;  DB 1;  Length 1173;
  Best Local Similarity   62.5%;  Pred. No. 4.4;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LRWPWWPW 9
            ::|||| |
Db   1112 IKWPWWVW 1119


RESULT   3
ADRO_BOVIN
ID   ADRO_BOVIN     STANDARD;      PRT;   492 AA.
AC   P08165;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NADPH:ADRENODOXIN OXIDOREDUCTASE, MITOCHONDRIAL PRECURSOR
DE   (EC 1.18.1.2) (ADRENODOXIN REDUCTASE) (AR) (FERREDOXIN-NADP(+)
DE   REDUCTASE).
GN   FDXR OR ADXR.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=94177140; PubMed=8130767;
RA   Takata Y., Sagara Y., Kono A., Sekimizu K., Horiuchi T.;
RT   "Gene structure of bovine adrenodoxin reductase.";
RL   Biol. Pharm. Bull. 16:1200-1206(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=88198050; PubMed=3448086;
RA   Sagara Y., Takata Y., Miyata T., Hara T., Horiuchi T.;
RT   "Cloning and sequence analysis of adrenodoxin reductase cDNA from
RT   bovine adrenal cortex.";
RL   J. Biochem. 102:1333-1336(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87270696; PubMed=3038094;
RA   Nonaka Y., Murakami H., Yabusaki Y., Kuramitsu S., Kagamiyama H.,
RA   Yamano T., Okamoto M.;
RT   "Molecular cloning and sequence analysis of full-length cDNA for mRNA
RT   of adrenodoxin oxidoreductase from bovine adrenal cortex.";
RL   Biochem. Biophys. Res. Commun. 145:1239-1247(1987).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal cortex;
RX   MEDLINE=89170752; PubMed=2924777;
RA   Hanukoglu I., Gutfinger T.;
```

```
RT   "cDNA sequence of adrenodoxin reductase. Identification of NADP-
RT   binding sites in oxidoreductases.";
RL   Eur. J. Biochem. 180:479-484(1989).
RN   [5]
RP   SEQUENCE OF N-TERMINUS, AND PARTIAL SEQUENCE.
RC   TISSUE=Adrenal cortex;
RX   MEDLINE=88082777; PubMed=3691502;
RA   Hanukoglu I., Gutfinger T., Haniu M., Shively J.E.;
RT   "Isolation of a cDNA for adrenodoxin reductase (ferredoxin-NADP+
RT   reductase). Implications for mitochondrial cytochrome P-450 systems.";
RL   Eur. J. Biochem. 169:449-455(1987).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 33-492.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=99299392; PubMed=10369776;
RA   Ziegler G.A., Vonrhein C., Hanukoglu I., Schulz G.E.;
RT   "The structure of adrenodoxin reductase of mitochondrial P450 systems:
RT   electron transfer for steroid biosynthesis.";
RL   J. Mol. Biol. 289:981-990(1999).
CC   -!- FUNCTION: SERVES AS THE FIRST ELECTRON TRANSFER PROTEIN IN ALL THE
CC       MITOCHONDRIAL P450 SYSTEMS. INCLUDING CHOLESTEROL SIDE CHAIN
CC       CLEAVAGE IN ALL STEROIDOGENIC TISSUES, STEROID 11-BETA
CC       HYDROXYLATION IN THE ADRENAL CORTEX, 25-OH-VITAMIN D3-24
CC       HYDROXYLATION IN THE KIDNEY, AND STEROL C-27 HYDROXYLATION IN THE
CC       LIVER.
CC   -!- CATALYTIC ACTIVITY: REDUCED ADRENODOXIN + NADP(+) = OXIDIZED
CC       ADRENODOXIN + NADPH.
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: CHOLESTEROL SIDE-CHAIN-CLEAVAGE SYSTEM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM (SHOWN HERE) AND A
CC       LONG FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE LONG FORM
CC       REPRESENTS 10-20% OF ALL ADRENODOXIN REDUCTASE MRNA. AND SEEMS TO
CC       BE INACTIVE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17029; AAA30362.1; -.
DR   EMBL; D00211; BAA00150.1; -.
DR   EMBL; X13736; CAA32002.1; -.
DR   PIR; A29604; A29604.
DR   PIR; JS0390; JS0390.
DR   PIR; S03558; S03558.
DR   PIR; JT0751; JT0751.
DR   PDB; 1CJC; 12-APR-99.
DR   PDB; 1E1L; 02-JUN-00.
DR   InterPro; IPR000759; Adrndx_redctse.
DR   PRINTS; PR00419; ADXRDTASE.
KW   Electron transport; Oxidoreductase; Flavoprotein; NADP; FAD;
KW   Mitochondrion; Transit peptide; Alternative splicing; 3D-structure.
FT   TRANSIT       1     32       MITOCHONDRION.
FT   CHAIN        33    492       NADPH:ADRENODOXIN OXIDOREDUCTASE.
FT   VARSPLIC    204    204       E -> EVLLLCQ (IN LONG ISOFORM).
FT   CONFLICT     77     77       G -> R (IN REF. 3).
FT   CONFLICT     81     94       FGVAPDHPEVKNVI -> VWLALTTPRSRMLL (IN REF.
FT                                3).
FT   CONFLICT    124    128       QDAYH -> RVYRLT (IN REF. 3).
FT   CONFLICT    268    268       K -> R (IN REF. 3).
FT   CONFLICT    317    318       PS -> RL (IN REF. 3).
FT   CONFLICT    323    333       RAAGIRLAVTR -> ARRSAWQSPE (IN REF. 3).
FT   CONFLICT    341    352       TRAVPTGDVEDL -> HPGSAHWGCGGP (IN REF. 3).
SQ   SEQUENCE   492 AA;  54338 MW;  E68F6F5D18F53131 CRC64;

  Query Match             57.0%;  Score 49;  DB 1;  Length 492;
  Best Local Similarity   83.3%;  Pred. No. 6.6;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 WPWWPW 9
           | ||||
Db       6 WRWWPW 11


RESULT   4
Y945_MYCTU
ID   Y945_MYCTU     STANDARD;      PRT;   253 AA.
AC   P71564;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE OXIDOREDUCTASE RV0945 (EC 1.-.-.-).
GN   RV0945 OR MT0971 OR MTCY10D7.29C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z79700; CAB02005.1; -.
DR   EMBL; AE006982; AAK45219.1; -.
DR   TIGR; MT0971; -.
DR   TubercuList; Rv0945; -.
DR   InterPro; IPR002198; ADH_short.
DR   Pfam; PF00106; adh_short; 1.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Hypothetical protein; Oxidoreductase; Complete proteome.
FT   ACT_SITE    159    159       BY SIMILARITY.
SQ   SEQUENCE   253 AA;  27138 MW;  BAD937208842DA12 CRC64;

  Query Match             54.7%;  Score 47;  DB 1;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 6.6;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       5 PWWPW 9
           |||||
Db     230 PWWPW 234


RESULT   5
YD55_MYCTU
ID   YD55_MYCTU     STANDARD;      PRT;   715 AA.
AC   Q11025;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 78.2 KDA PROTEIN RV1355C.
GN   RV1355C OR MT1398 OR MTCY02B10.19C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z75555; CAA99988.1; -.
DR   EMBL; AE007012; AAK45661.1; ALT_INIT.
DR   TIGR; MT1398; -.
DR   TubercuList; Rv1355c; -.
DR   InterPro; IPR000594; ThiF_family.
DR   Pfam; PF00899; ThiF_family; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   715 AA;  78181 MW;  455495248A56041C CRC64;


  Query Match             54.7%;  Score 47;  DB 1;  Length 715;
  Best Local Similarity   66.7%;  Pred. No. 17;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 RWPWWPWRR 11
           || ::||||
Db      65 RWAYYPWRR 73


RESULT   6
```

```
TRPE_PSESS
ID   TRPE_PSESS     STANDARD;      PRT;   505 AA.
AC   P21689;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN   TRPE.
OS   Pseudomonas syringae (pv. savastanoi).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=29438;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91100331; PubMed=1987141;
RA   da Costa E., Silva O., Kosuge T.;
RT   "Molecular characterization and expression analysis of the
RT   anthranilate synthase gene of Pseudomonas syringae subsp.
RT   savastanoi.";
RL   J. Bacteriol. 173:463-471(1991).
CC   -!- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
CC       PYRUVATE + L-GLUTAMATE.
CC   -!- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC   -!- SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II (BY
CC       SIMILARITY).
CC   -!- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
CC       USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
CC       GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55911; AAA26016.1; -.
DR   PIR; A39128; A39128.
DR   InterPro; IPR000350; Chorismate_bind.
DR   Pfam; PF00425; chorismate_bind; 1.
DR   PRINTS; PR00095; ANTSNTHASEI.
DR   ProDom; PD000779; Chorismate_bind; 1.
KW   Tryptophan biosynthesis; Lyase.
SQ   SEQUENCE   505 AA;  56084 MW;  A38E81931331F6BB CRC64;


  Query Match             52.9%;  Score 45.5;  DB 1;  Length 505;
  Best Local Similarity   28.0%;  Pred. No. 19;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels   13;  Gaps    1;

Qy       1 ILRW----------PWWPWRRK 12
             :| |              |||| |:
Db     467 VLEWEETLNKRRAMVGSAWWPWPRR 491


RESULT   7
YA05_SCHPO
ID   YA05_SCHPO     STANDARD;      PRT;   196 AA.
AC   Q09677;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL 22.1 KDA PROTEIN C5H10.05C IN CHROMOSOME I.
GN   SPAC5H10.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO BACTERIAL MODULATOR OF DRUG ACTIVITY B
CC       (MDAB).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49811; CAA89955.1; -.
DR   InterPro; IPR003680; NADHdh_2.
DR   Pfam; PF02525; NADHdh_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   196 AA;  22104 MW;  436764DA9E26074C CRC64;


  Query Match             52.3%;  Score 45;  DB 1;  Length 196;
  Best Local Similarity   50.0%;  Pred. No. 9.5;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    4;  Gaps    2;

Qy       1 ILRWP-WW---PWRRK 12
           | :|| ||   ||: |
Db      63 IYQWPGWWMGTPWKLK 78


RESULT   8
CN3B_RAT
ID   CN3B_RAT       STANDARD;      PRT;  1108 AA.
AC   Q63085;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CGMP-INHIBITED 3',5'-CYCLIC PHOSPHODIESTERASE B (EC 3.1.4.17) (CYCLIC
DE   GMP INHIBITED PHOSPHODIESTERASE B) (CGI-PDE B) (CGIPDE1).
GN   PDE3B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Adipose tissue;
RX   MEDLINE=93366761; PubMed=8395509;
RA   Taira M., Hockman S.C., Calvo J.C., Taira M., Belfrage P.,
RA   Manganiello V.C.;
RT   "Molecular cloning of the rat adipocyte hormone-sensitive cyclic GMP-
RT   inhibited cyclic nucleotide phosphodiesterase.";
RL   J. Biol. Chem. 268:18573-18579(1993).
CC   -!- FUNCTION: MAY PLAY A ROLE IN FAT METABOLISM.
CC   -!- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
CC       GUANOSINE 5'-PHOSPHATE.
CC   -!- ENZYME REGULATION: INHIBITED BY CGMP.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: ABUNDANT IN ADIPOSE TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z22867; CAA80489.1; -.
DR   InterPro; IPR003607; HDc.
DR   InterPro; IPR002073; PDEase.
DR   Pfam; PF00233; PDEase; 1.
DR   SMART; SM00471; HDc; 1.
DR   PROSITE; PS00126; PDEASE_I; 1.
KW   Hydrolase; cGMP; Membrane.
FT   DOMAIN       16     22       POLY-PRO.
FT   DOMAIN       99    102       POLY-ALA.
FT   DOMAIN      175    179       POLY-ALA.
FT   DOMAIN     1007   1021       POLY-ASP.
FT   DOMAIN     1068   1071       POLY-GLU.
FT   DOMAIN     1101   1104       POLY-GLU.
SQ   SEQUENCE   1108 AA;  123105 MW;  C9B5078C7D3ADD6D CRC64;


  Query Match             52.3%;  Score 45;  DB 1;  Length 1108;
  Best Local Similarity   62.5%;  Pred. No. 46;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 WPWWPWRR 11
           | || | |
Db     164 WQWWSWLR 171


RESULT   9
FUT3_HUMAN
ID   FUT3_HUMAN     STANDARD;      PRT;   361 AA.
AC   P21217; Q99448; Q99449;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GALACTOSIDE 3(4)-L-FUCOSYLTRANSFERASE (EC 2.4.1.65) (BLOOD GROUP LEWIS
DE   ALPHA-4-FUCOSYLTRANSFERASE) (LEWIS FT) (FUCOSYLTRANSFERASE 3) (FUCT-
DE   III).
GN   FUT3 OR LE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91032981; PubMed=1977660;
RA   Kukowska-Latallo J.F., Larsen R.D., Nair R.P., Lowe J.B.;
RT   "A cloned human cDNA determines expression of a mouse stage-specific
RT   embryonic antigen and the Lewis blood group
RT   alpha(1,3/1,4)fucosyltransferase.";
RL   Genes Dev. 4:1288-1303(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95378269; PubMed=7650030;
RA   Cameron H.S., Szczepaniak D., Weston W.;
RT   "Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
RT   genes in normal tissues. Alternative splicing, polyadenylation, and
RT   isoforms.";
RL   J. Biol. Chem. 270:20112-20122(1995).
RN   [3]
RP   VARIANT LE(-) MET-105.
RX   MEDLINE=94059067; PubMed=8240322;
RA   Elmgren A., Rydberg L., Larson G.;
RT   "Genotypic heterogeneity among Lewis negative individuals.";
RL   Biochem. Biophys. Res. Commun. 196:515-520(1993).
RN   [4]
RP   VARIANTS LE(-) ARG-20; SER-170 AND ALA-336.
RX   MEDLINE=94059082; PubMed=8240337;
RA   Nishihara S., Yazawa S., Iwasaki H., Nakazato M., Kudo T., Ando T.,
RA   Narimatsu H.;
RT   "Alpha (1,3/1,4)fucosyltransferase (FucT-III) gene is inactivated by
RT   a single amino acid substitution in Lewis histo-blood type negative
RT   individuals.";
RL   Biochem. Biophys. Res. Commun. 196:624-631(1993).
RN   [5]
RP   VARIANTS LE(-) ARG-20 AND SER-170.
RX   MEDLINE=94033579; PubMed=8219240;
```

```
RA   Koda Y., Kimura H., Mekada E.;
RT   "Analysis of Lewis fucosyltransferase genes from the human gastric
RT   mucosa of Lewis-positive and -negative individuals.";
RL   Blood 82:2915-2919(1993).
RN   [6]
RP   VARIANTS LE(-) ARG-20 AND LYS-356.
RX   MEDLINE=94342259; PubMed=8063716;
RA   Mollicone R., Reguigne I., Kelly R.J., Fletcher A., Watt J.,
RA   Chatfield S., Aziz A., Cameron H.S., Weston B.W., Lowe J.B., Oriol R.;
RT   "Molecular basis for Lewis alpha(1,3/1,4)-fucosyltransferase gene
RT   deficiency (FUT3) found in Lewis-negative Indonesian pedigrees.";
RL   J. Biol. Chem. 269:20987-20994(1994).
RN   [7]
RP   VARIANT LE(-) LYS-356.
RX   MEDLINE=95050753; PubMed=7961897;
RA   Nishihara S., Narimatsu H., Iwasaki H., Yazawa S., Akamatsu S.,
RA   Ando T., SenO T., Narimatsu I.;
RT   "Molecular genetic analysis of the human Lewis histo-blood group
RT   system.";
RL   J. Biol. Chem. 269:29271-29278(1994).
RN   [8]
RP   VARIANTS LE(-) ARG-20; ARG-68; MET-105 AND LYS-356.
RX   MEDLINE=96243526; PubMed=8801770;
RA   Elmgren A., Boerjeson C., Svensson L., Rydberg L., Larson G.;
RT   "DNA sequencing and screening for point mutations in the human Lewis
RT   'FUT3' gene enables molecular genotyping of the human Lewis blood
RT   group system.";
RL   Vox Sang. 70:97-103(1996).
RN   [9]
RP   VARIANTS LE(-) ARG-68 AND MET-105.
RX   MEDLINE=97413801; PubMed=9268337;
RA   Elmgren A., Mollicone R., Costache M., Boerjeson C., Oriol R.,
RA   Harrington J., Larson G.;
RT   "Significance of individual point mutations, T202C and C314T, in the
RT   human Lewis 'FUT3' gene for expression of Lewis antigens by the human
RT   alpha'1,3/1,4'-fucosyltransferase, Fuc-TIII.";
RL   J. Biol. Chem. 272:21994-21998(1997).
RN   [10]
RP   VARIANTS LE(+) K-102; A-124, AND VARIANTS LE(-) N-162; R-223; M-270.
RX   MEDLINE=98366989; PubMed=9703429;
RA   Pang H., Liu Y., Koda Y., Soejima M., Jia J., Schlaphoff T.,
RA   du Toit E.D., Kimura H.;
RT   "Five novel missense mutations of the Lewis gene 'FUT3' in African
RT   'Xhosa' and Caucasian populations in South Africa.";
RL   Hum. Genet. 102:675-680(1998).
CC   -!- FUNCTION: MAY CATALYZE ALPHA-1,3 AND ALPHA-1,4 GLYCOSIDIC LINKAGES
CC       INVOLVED IN THE EXPRESSION OF VIM-2, LEWIS A, LEWIS B, SIALYL
CC       LEWIS X AND LEWIS X/SSEA-1 ANTIGENS. MAY BE INVOLVED IN BLOOD
CC       GROUP LEWIS DETERMINATION; LEWIS-POSITIVE (LE(+)) INDIVIDUALS
CC       HAVE AN ACTIVE ENZYME WHILE LEWIS-NEGATIVE (LE(-)) INDIVIDUALS
CC       HAVE AN INACTIVE ENZYME.
CC   -!- CATALYTIC ACTIVITY: GDP-L-FUCOSE + 1,3-BETA-D-GALACTOSYL-
CC       N-ACETYL-D-GLUCOSAMINYL-R = GDP + 1,3-BETA-D-GALACTOSYL-
CC       (ALPHA-1,4-L-FUCOSYL)-N-ACETYL-D-GLUCOSAMINYL-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN STOMACH, COLON, SMALL
CC       INTESTINE, LUNG AND KIDNEY AND TO A LESSER EXTENT IN SALIVARY
CC       GLAND, BLADDER, UTERUS AND LIVER.
CC   -!- MISCELLANEOUS: ALSO ACTS ON THE CORRESPONDING 1,4-GALACTOSYL
CC       DERIVATIVE, FORMING 1,3-L-FUCOSYL LINKS.
CC   -!- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
CC       GLYCOSYLTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53578; CAA37641.1; -.
DR   EMBL; U27328; AAC50187.1; -.
DR   EMBL; U27326; AAC50185.1; -.
DR   EMBL; U27327; AAC50186.1; -.
DR   EMBL; D89324; BAA13941.1; -.
DR   EMBL; D89325; BAA13942.1; -.
DR   PIR; A36669; A36669.
DR   MIM; 111100; -.
DR   InterPro; IPR001503; Glyco_transf_10.
DR   Pfam; PF00852; Glyco_transf_10; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Polymorphism; Blood group antigen.
FT   DOMAIN        1     15       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     16     34       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       35    361       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    154    154       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   CARBOHYD    185    185       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   VARIANT      20     20       L -> R (IN LE(-)).
FT                                /FTId=VAR_003426.
FT   VARIANT      68     68       W -> R (IN LE(-)).
FT                                /FTId=VAR_007959.
FT   VARIANT     102    102       Q -> K (IN LE(+)).
FT                                /FTId=VAR_007960.
FT   VARIANT     105    105       T -> M (IN LE(-)).
FT                                /FTId=VAR_003427.
FT   VARIANT     124    124       S -> A (IN LE(+)).
FT                                /FTId=VAR_007961.
FT   VARIANT     162    162       D -> N (IN LE(-)).
FT                                /FTId=VAR_007962.
FT   VARIANT     170    170       G -> S (IN LE(-); COMPLETELY INACTIVE).
FT                                /FTId=VAR_003428.
FT   VARIANT     223    223       G -> R (IN LE(-)).
FT                                /FTId=VAR_007963.
FT   VARIANT     270    270       V -> M (IN LE(-)).
FT                                /FTId=VAR_007964.
FT   VARIANT     336    336       D -> A (IN LE(-)).
FT                                /FTId=VAR_003429.
FT   VARIANT     356    356       I -> K (IN LE(-); LESS THAN 10% REDUCTION
FT                                IN ACTIVITY).
FT                                /FTId=VAR_003430.
SQ   SEQUENCE   361 AA;  42117 MW;  BF4398044F19C284 CRC64;


  Query Match             51.2%;  Score 44;  DB 1;  Length 361;
  Best Local Similarity   85.7%;  Pred. No. 23;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PWWPWRR 11
            | |||||
Db       9 PQWPWRR 15


RESULT  10
FUT3_PANTR
ID   FUT3_PANTR     STANDARD;      PRT;   372 AA.
AC   O19058;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GALACTOSIDE 3(4)-L-FUCOSYLTRANSFERASE (EC 2.4.1.65) (BLOOD GROUP LEWIS
DE   ALPHA-4-FUCOSYLTRANSFERASE) (LEWIS FT) (FUCOSYLTRANSFERASE 3) (FUCT-
DE   III) (ALPHA-3/4-FUCOSYLTRANSFERASE).
GN   FUT3.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98037800; PubMed=9368041;
RA   Costache M., Apoil P.-A., Cailleau A., Elmgren A., Larson G.,
RA   Henry S., Blancher A., Iordachescu D., Oriol R., Mollicone R.;
RT   "Evolution of fucosyltransferase genes in vertebrates.";
```

```
RL   J. Biol. Chem. 272:29721-29728(1997).
CC   -!- FUNCTION: MAY CATALYZE ALPHA-1,3 AND ALPHA-1,4 GLYCOSIDIC LINKAGES
CC       INVOLVED IN THE EXPRESSION OF SIALYL LEWIS X AND LEWIS X/SSEA-1
CC       ANTIGENS. IT MAY BE INVOLVED IN BLOOD GROUP LEWIS DETERMINATION
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: GDP-L-FUCOSE + 1,3-BETA-D-GALACTOSYL-
CC       N-ACETYL-D-GLUCOSAMINYL-R = GDP + 1,3-BETA-D-GALACTOSYL-
CC       (ALPHA-1,4-L-FUCOSYL)-N-ACETYL-D-GLUCOSAMINYL-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC       FORM IN TRANS CISTERNAE OF GOLGI (BY SIMILARITY).
CC   -!- POLYMORPHISM: THERE ARE TWO ALLELES (A AND B). ALLELE A HAS ARG-
CC       162 AND VAL-304. ALLELE B HAS GLY-162 AND MET-304.
CC   -!- MISCELLANEOUS: ALSO ACTS ON THE CORRESPONDING 1,4-GALACTOSYL
CC       DERIVATIVE, FORMING 1,3-L-FUCOSYL LINKS.
CC   -!- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
CC       GLYCOSYLTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14033; CAA74360.1; -.
DR   InterPro; IPR001503; Glyco_transf_10.
DR   Pfam; PF00852; Glyco_transf_10; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Polymorphism.
FT   DOMAIN        1     14       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     15     34       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       35    372       LUMENAL, CATALYTIC (POTENTIAL)..
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     162    162       R -> G (IN ALLELE B).
FT   VARIANT     304    304       V -> M (IN ALLELE B).
SQ   SEQUENCE   372 AA;  43233 MW;  649CBF8BCA7BD74C CRC64;


  Query Match             51.2%;  Score 44;  DB 1;  Length 372;
  Best Local Similarity   85.7%;  Pred. No. 23;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PWWPWRR 11
            | |||||
Db       9 PQWPWRR 15


RESULT  11
MML6_MYCTU
ID   MML6_MYCTU     STANDARD;      PRT;   397 AA.
AC   Q10773;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN MMPL6.
GN   MMPL6 OR RV1557 OR MT1608 OR MTCY48.08C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74020; CAA98334.1; -.
DR   EMBL; AE007027; AAK45875.1; -.
DR   TIGR; MT1608; -.
DR   TubercuList; Rv1557; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM    161    181       POTENTIAL.
FT   TRANSMEM    190    210       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    242    262       POTENTIAL.
FT   TRANSMEM    293    313       POTENTIAL.
FT   TRANSMEM    330    350       POTENTIAL.
SQ   SEQUENCE   397 AA;  42421 MW;  678DC86E24472BF4 CRC64;


  Query Match             51.2%;  Score 44;  DB 1;  Length 397;
  Best Local Similarity   75.0%;  Pred. No. 25;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 RWPWWPWR 10
            || ||| |
Db     351 RWFWWPQR 358


RESULT  12
YDW6_SCHPO
ID   YDW6_SCHPO     STANDARD;      PRT;   535 AA.
AC   O13912;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL 60.1 KDA PROTEIN C23C11.06C IN CHROMOSOME I.
GN   SPAC23C11.06C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z98559; CAB11159.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      55     75       POTENTIAL.
FT   TRANSMEM      82    102       POTENTIAL.
FT   TRANSMEM     115    135       POTENTIAL.
FT   TRANSMEM     143    163       POTENTIAL.
FT   TRANSMEM     201    221       POTENTIAL.
FT   TRANSMEM     346    366       POTENTIAL.
SQ   SEQUENCE   535 AA;  60124 MW;  A6AE149AAA2929E2 CRC64;
```

```
  Query Match             51.2%;  Score 44;  DB 1;  Length 535;
  Best Local Similarity   50.0%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    4;  Gaps    1;

Qy       4 WPW----WPWRR 15
           | |    ||||:
Db     183 WSWSPSTWPWRQ 194
```

```
RESULT  13
MML4_MYCTU
ID   MML4_MYCTU     STANDARD;      PRT;   967 AA.
AC   O53735;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN MMPL4.
GN   MMPL4 OR RV0450C OR MT0466 OR MTV037.14C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL021932; CAA17407.1; -.
DR   EMBL; AE006949; AAK44689.1; -.
DR   TIGR; MT0466; -.
DR   TubercuList; Rv0450c; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      26     46       POTENTIAL.
FT   TRANSMEM     210    230       POTENTIAL.
FT   TRANSMEM     242    262       POTENTIAL.
FT   TRANSMEM     303    323       POTENTIAL.
FT   TRANSMEM     333    353       POTENTIAL.
FT   TRANSMEM     384    404       POTENTIAL.
FT   TRANSMEM     769    789       POTENTIAL.
FT   TRANSMEM     793    813       POTENTIAL.
FT   TRANSMEM     821    841       POTENTIAL.
FT   TRANSMEM     875    895       POTENTIAL.
FT   TRANSMEM     896    916       POTENTIAL.
SQ   SEQUENCE   967 AA;  105234 MW;  6301014031480484 CRC64;
```

```
  Query Match             51.2%;  Score 44;  DB 1;  Length 967;
  Best Local Similarity   75.0%;  Pred. No. 55;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 RWPWWPWR 10
           || ||| |
Db     930 RWFWWPLR 937
```

```
RESULT  14
MML2_MYCTU
ID   MML2_MYCTU     STANDARD;      PRT;   968 AA.
AC   Q11171;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN MMPL2.
GN   MMPL2 OR RV0507 OR MT0528 OR MTCY20G9.34.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z77162; CAB00933.1;. -.
DR   EMBL; AE006953; AAK44751.1; -.
DR   TIGR; MT0528; -.
DR   TubercuList; Rv0507; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     22     42       POTENTIAL.
FT   TRANSMEM    204    224       POTENTIAL.
FT   TRANSMEM    245    265       POTENTIAL.
FT   TRANSMEM    297    317       POTENTIAL.
FT   TRANSMEM    328    348       POTENTIAL.
FT   TRANSMEM    378    398       POTENTIAL.
FT   TRANSMEM    763    783       POTENTIAL.
FT   TRANSMEM    787    807       POTENTIAL.
FT   TRANSMEM    815    835       POTENTIAL.
FT   TRANSMEM    866    886       POTENTIAL.
FT   TRANSMEM    891    911       POTENTIAL.
FT   CONFLICT    426    426       R -> H (IN REF. 2).
FT   CONFLICT    656    656       E -> A (IN REF. 2).
SQ   SEQUENCE   968 AA;  106201 MW;  B68AE9B78164EDC0 CRC64;


  Query Match             51.2%;  Score 44;  DB 1;  Length 968;
  Best Local Similarity   75.0%;  Pred. No. 55;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 RWPWWPWR 10
          || ||| |
Db    924 RWFWWPLR 931


RESULT  15
SX13_MOUSE
ID   SX13_MOUSE     STANDARD;      PRT;   984 AA.
AC   Q04891;
DT   01-JUN-1994 (Rel. 29, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SOX-13 PROTEIN.
GN   SOX13 OR SOX-13.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND FUNCTION (ISOFORM 1).
RC   TISSUE=Embryo;
RX   MEDLINE=98083175; PubMed=9421502;
RA   Roose J., Korver W., Oving E., Wilson A., Wagenaar G., Markman M.,
RA   Lamers W., Clevers H.;
RT   "High expression of the HMG box factor sox-13 in arterial walls during
RT   embryonic development.";
RL   Nucleic Acids Res. 26:469-476(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND FUNCTION (ISOFORM 2).
RC   TISSUE=Embryo;
RX   MEDLINE=98201614; PubMed=9524265;
RA   Kido S., Hiraoka Y., Ogawa M., Sakai Y., Yoshimura Y., Aiso S.;
RT   "Cloning and characterization of mouse mSox13 cDNA.";
RL   Gene 208:201-206(1998).
RN   [3]
RP   SEQUENCE OF 405-460 FROM N.A.
RX   MEDLINE=93181275; PubMed=8441686;
RA   Wright E.M., Snopek B., Koopman P.;
RT   "Seven new members of the Sox gene family expressed during mouse
RT   development.";
RL   Nucleic Acids Res. 21:744-744(1993).
CC   -!- FUNCTION: BINDS TO THE SEQUENCE 5'-AACAAT-3'.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ISOFORM 1 (SHOWN HERE) AND
CC       ISOFORM 2; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, HIGH LEVELS OF EXPRESSION ARE
CC       FOUND IN THE ARTERIAL WALLS AT 13.5 DAYS POST COITUM (DPC). LOW
CC       LEVELS ARE FOUND IN THE INNER EAR AT 13.5 DPC AND IN SOME CELLS IN
CC       THE THYMUS AT 16.5 DPC. EXPRESSED IN THE TRACHEAL EPITHELIUM BELOW
CC       THE VOCAL CORD AND IN THE HAIR FOLLICLES AT 18 DPC.
CC   -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ000740; CAA04278.1; -.
DR   EMBL; AB006329; BAA25786.1; -.
DR   EMBL; Z18962; CAA79487.1; -.
DR   PIR; S30241; S30241.
DR   HSSP; Q05066; 1HRZ.
DR   MGD; MGI:98361; Sox13.
DR   InterPro; IPR000910; HMG_12_box.
DR   Pfam; PF00505; HMG_box; 2.
DR   SMART; SM00398; HMG; 1.
KW   DNA-binding; Nuclear protein; Alternative splicing.
FT   DOMAIN      159    195       GLN-RICH.
FT   DNA_BIND    397    465       HMG BOX.
FT   VARSPLIC    495    519       PGCSPKLHHPVSRPSLVARGGLWLL -> QGARQSYTIP
FT                                (IN ISOFORM 2).
FT   VARSPLIC    603    609       SWWYSQT -> ELVVLTD (IN ISOFORM 2).
FT   VARSPLIC    610    984       MISSING (IN ISOFORM 2).
FT   CONFLICT     35     35       P -> L (IN REF. 2).
FT   CONFLICT     41     42       AT -> TN (IN REF. 2).
FT   CONFLICT    195    195       Q -> QQ (IN REF. 2).
SQ   SEQUENCE   984 AA;  108897 MW;  7F5506EDADEB98C5 CRC64;


  Query Match             51.2%;  Score 44;  DB 1;  Length 984;
  Best Local Similarity   42.9%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    8;  Gaps    1;

Qy      4 WPWW--------PW 9
          ||||        ||
Db    686 WPWWTKLAEGFSPW 699


Search completed: January  4, 2002, 08:47:48
Job time: 406 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 4, 2002, 08:41:32 ; Search time 27.18 Seconds
(without alignments)
33.631 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRWPWWPWRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	70	81.4	144	1	JC1222	indolicidin precur
2	53	61.6	1173	1	VGIHHC	E2 glycoprotein pr
3	51	59.3	299	2	T12505	hypothetical prote
4	50	58.1	111	2	T29295	hypothetical prote
5	49	57.0	498	1	JT0751	ferredoxin--NADP+
6	48.5	56.4	114	2	T36208	hypothetical prote
7	47	54.7	248	2	S23449	NADH oxidase (H2O2
8	47	54.7	253	2	G70715	hypothetical prote
9	47	54.7	276	2	B83161	probable short-cha
10	47	54.7	715	2	B70741	probable moeY prot
11	47	54.7	1411	2	T48529	hypothetical prote
12	46	53.5	728	2	T51071	related to trfA pr
13	45.5	52.9	505	2	A39128	anthranilate synth
14	45	52.3	196	2	S55483	modulator of drug
15	45	52.3	273	2	F82646	monofunctional bio
16	45	52.3	412	2	A83604	probable MFS trans
17	45	52.3	448	2	H72376	hypothetical prote
18	45	52.3	1108	2	A48508	cyclic-nucleotide
19	44	51.2	257	2	S70177	yfrE protein - Yer
20	44	51.2	361	2	A36669	galactoside 3(4)-L
21	44	51.2	397	2	B70763	probable membrane
22	44	51.2	535	2	T38244	hypothetical prote
23	44	51.2	621	2	S37664	peplomeric polypro
24	44	51.2	630	2	S37663	peplomeric polypro
25	44	51.2	967	2	C70831	probable mmpL4 pro
26	44	51.2	968	2	F70746	probable mmpL2 pro
27	44	51.2	968	2	T00322	hypothetical prote
28	44	51.2	1154	1	VGIHIB	E2 glycoprotein pr
29	44	51.2	1162	1	VGIHAK	E2 glycoprotein pr
30	44	51.2	1162	2	S07421	E2 glycoprotein pr
31	44	51.2	1162	2	S14939	E2 glycoprotein pr
32	44	51.2	1162	2	S14940	E2 glycoprotein pr
33	43.5	50.6	1529	2	A59189	ATP-binding casset
34	43	50.0	51	2	S23291	light-harvesting p
35	43	50.0	192	2	H86543	hypothetical prote
36	43	50.0	192	2	D72081	conserved hypothet
37	43	50.0	236	2	JQ0606	arylesterase (EC 3
38	43	50.0	250	2	A83506	probable cobalamin
39	43	50.0	278	2	T46458	hypothetical prote
40	43	50.0	298	2	B72492	hypothetical prote
41	43	50.0	646	2	H82555	c-type cytochrome
42	43	50.0	711	2	C40046	antibiotic transpo
43	43	50.0	738	2	F96701	hypothetical prote
44	43	50.0	958	2	A70634	probable mmpL1 pro
45	43	50.0	1112	2	S70522	cyclic nucleotide

ALIGNMENTS

RESULT 1
JC1222
indolicidin precursor - bovine
N;Alternate names: antimicrobial peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JC1222; A42387; S25664
R;del Sal, G.; Storici, P.; Schneider, C.; Romeo, D.; Zanetti, M.
Biochem. Biophys. Res. Commun. 187, 467-472, 1992
A;Title: cDNA cloning of the neutrophil bactericidal peptide indolicidin.
A;Reference number: JC1222; MUID:92392368
A;Accession: JC1222
A;Molecule type: mRNA
A;Residues: 1-144 <SAL>
A;Cross-references: EMBL:X67340; NID:g462; PIDN:CAA47755.1; PID:g463
A;Experimental source: bone marrow
R;Selsted, M.E.; Novotny, M.J.; Morris, W.L.; Tang, Y.Q.; Smith, W.; Cullor, J.S.
J. Biol. Chem. 267, 4292-4295, 1992
A;Title: Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils.
A;Reference number: A42387; MUID:92165771
A;Accession: A42387
A;Molecule type: protein
A;Residues: 131-143 <SEL>
A;Experimental source: neutrophils
A;Note: sequence extracted from NCBI backbone (NCBIP:83840)
C;Superfamily: cathelin; cystatin homology
C;Keywords: amidated carboxyl end
F;1-29/Domain: signal sequence #status predicted <SIG>
F;22-129/Domain: cystatin homology <CYS>
F;30-130/Domain: propeptide #status predicted <PRO>
F;131-143/Product: indolicidin #status experimental <MAT>
F;143/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 81.4%; Score 70; DB 1; Length 144;
Best Local Similarity 88.9%; Pred. No. 0.011;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 RWPWWPWRR 11
           :|||||||| 
Db     135 KWPWWPWRR 143
```

RESULT 2
VGIHHC
E2 glycoprotein precursor - human coronavirus (strain 229E)
N;Alternate names: peplomer glycoprotein; S glycoprotein; spike glycoprotein
C;Species: human coronavirus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C;Accession: A34766; S05460
R;Raabe, T.; Schelle-Prinz, B.; Siddell, S.G.

J. Gen. Virol. 71, 1065-1073, 1990
A;Title: Nucleotide sequence of the gene encoding the spike glycoprotein of human corona
A;Reference number: A34766; MUID:90264837
A;Accession: A34766
A;Molecule type: mRNA
A;Residues: 1-1173 <RAA>
A;Cross-references: EMBL:X16816; NID:g58926; PIDN:CAA34723.1; PID:g58927
A;Experimental source: strain 229E
R;Raabe, T.; Siddell, S.
Nucleic Acids Res. 17, 6387, 1989
A;Title: Nucleotide sequence of the human coronavirus HCV 229E mRNA 4 and mRNA 5 unique
A;Reference number: A34038; MUID:89366667
A;Accession: S05460
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1159-1173 <RA2>
A;Cross-references: EMBL:X15654; NID:g58921; PIDN:CAA33680.1; PID:g1334827
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-1173/Product: E2 glycoprotein #status predicted <MAT>
F;1116-1138/Domain: transmembrane #status predicted <TMN>
F;23,62,98,147,171,176,220,243,326,333,440,464,518,538,542,568,581,587,663,671,930,1015,

  Query Match 61.6%; Score 53; DB 1; Length 1173;
  Best Local Similarity 62.5%; Pred. No. 12;
  Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LRWPWWPW 9
::|||| |
Db 1112 IKWPWWVW 1119

RESULT 3
T12505
hypothetical protein DKFZp434C192.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C;Accession: T12505
R;Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z17527
A;Accession: T12505
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-299 <ANS>
A;Cross-references: EMBL:AL096753
A;Experimental source: adult testis; clone DKFZp434C192
C;Genetics:
A;Note: DKFZp434C192.1

  Query Match 59.3%; Score 51; DB 2; Length 299;
  Best Local Similarity 85.7%; Pred. No. 6.1;
  Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PWWPWRR 11
|||||
Db 37 PWWPWSR 43

RESULT 4
T29295
hypothetical protein C50F7.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29295
R;Johnson, D.; Stellyes, L.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C50F7.
A;Reference number: Z20601
A;Accession: T29295
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-111 <JOH>
A;Cross-references: EMBL:U41557; PIDN:AAA83303.1; CESP:C50F7.8
C;Genetics:
A;Gene: CESP:C50F7.8

  Query Match 58.1%; Score 50; DB 2; Length 111;
  Best Local Similarity 54.5%; Pred. No. 3.2;
  Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ILRWPWWPWRR 11
:: ||||| |
Db 12 VMWWPWWPGGR 22

RESULT 5
JT0751
ferredoxin--NADP+ reductase (EC 1.18.1.2), long form precursor - bovine
N;Alternate names: adrenodoxin reductase
C;Species: Bos primigenius taurus (cattle)
C;Date: 14-Jul-1994 #sequence_revision 18-Oct-1996 #text_change 16-Jun-2000
C;Accession: JT0751; JT0079; JS0390; S03558; PS0003; A29604; S52100
R;Takata, Y.; Sagara, Y.; Kono, A.; Sekimizu, K.; Horiuchi, T.
Biol. Pharm. Bull. 16, 1200-1206, 1993
A;Title: Gene structure of bovine adrenodoxin reductase.
A;Reference number: JT0751; MUID:94177140
A;Accession: JT0751
A;Molecule type: DNA
A;Residues: 1-498 <TAK>
A;Cross-references: GB:D83475; NID:g1199916; PIDN:BAA11921.1; PID:g4521308
A;Experimental source: adrenal cortex
A;Note: the authors translated the codon GTC for residue 205 as Gly
R;Sagara, Y.; Takata, Y.; Miyata, T.; Hara, T.; Horiuchi, T.
J. Biochem. 102, 1333-1336, 1987
A;Title: Cloning and sequence analysis of adrenodoxin reductase cDNA from bovine adre
A;Reference number: JT0079; MUID:88198050
A;Accession: JT0079
A;Molecule type: mRNA
A;Residues: 1-204,211-498 <SAG>
A;Cross-references: GB:D00211; NID:g217433; PIDN:BAA00150.1; PID:g217434
A;Note: the deduced sequence is partially confirmed by amino acid sequencing of 15 is
R;Sagara, Y.
submitted to DDBJ, September 1989
A;Reference number: JS0390
A;Contents: revision, insertion of residues 205-210
A;Accession: JS0390
A;Molecule type: mRNA
A;Residues: 56-498 <SA2>
R;Hanukoglu, I.; Gutfinger, T.
Eur. J. Biochem. 180, 479-484, 1989
A;Title: cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites
A;Reference number: S03558; MUID:89170752
A;Accession: S03558
A;Molecule type: mRNA
A;Residues: 155-204,211-498 <HAN>
A;Cross-references: EMBL:X13736; NID:g65; PIDN:CAA32002.1; PID:g833776
A;Note: 405-Ser was also found
R;Hamamoto, I.; Kurokohchi, K.; Tanaka, S.; Ichikawa, Y.
Biochim. Biophys. Acta 953, 207-213, 1988
A;Title: Adrenoferredoxin-binding peptide of NADPH-adrenoferredoxin reductase.
A;Reference number: PS0003; MUID:88184054
A;Accession: PS0003
A;Molecule type: protein
A;Residues: 33-41,'S',43-62;260-283,'TM';496-498 <HAM>
A;Note: a cyanogen bromide peptide binds to adrenoferredoxin
R;Nonaka, Y.; Murakami, H.; Yabusaki, Y.; Kuramitsu, S.; Kagamiyama, H.; Yamano, T.;
Biochem. Biophys. Res. Commun. 145, 1239-1247, 1987
A;Title: Molecular cloning and sequence analysis of full-length cDNA for mRNA of adre
A;Reference number: A29604; MUID:87270696
A;Accession: A29604
A;Molecule type: mRNA

A;Residues: 1-76,'R',78-80,'VWLALTTPRSRMLL',95-123,'RVYRLT',129-204,211-273,'R',275-322,
A;Cross-references: GB:M17029; NID:g162628; PIDN:AAA30362.1; PID:g162629
A;Experimental source: adrenal cortex
R;Warburton, R.J.; Seybert, D.W.
Biochim. Biophys. Acta 1246, 39-46, 1995
A;Title: Structural and functional characterization of bovine adrenodoxin reductase by l
A;Reference number: S52100; MUID:95110846
A;Accession: S52100
A;Status: preliminary
A;Molecule type: protein
A;Residues: 'X',34-41,'X',43-48,'X',50-51;304-306,'X',308-309,'X',311-326 <WAR>
C;Comment: Ferredoxin--NADP+ reductase is localized in the matrix of adrenal cortex mito
ferredoxin--NADP+ reductase, adrenodoxin and two forms of cytochrome P-450.
C;Genetics:
A;Introns: 27/1; 59/3; 91/3; 132/3; 170/3; 204/3; 246/3; 275/1; 341/3; 399/1; 456/1
C;Function:
A;Description: catalyzes the reversible reduction of NADP+ by reduced ferredoxin or redu
C;Superfamily: human ferredoxin--NADP+ reductase
C;Keywords: alternative splicing; flavoprotein; mitochondrion; monomer; NADP; oxidoreduc
F;1-32/Domain: transit peptide (mitochondrion) #status predicted <SIG>
F;33-498/Product: ferredoxin--NADP+ reductase, long form #status predicted <MAT>
F;33-204,211-498/Product: ferredoxin--NADP+ reductase, short form #status experimental <
F;40-70/Region: beta-alpha-beta FAD nucleotide-binding fold
F;180-190/Region: NADP binding #status predicted
F;281/Binding site: substrate (Lys) #status experimental

  Query Match 57.0%; Score 49; DB 1; Length 498;
  Best Local Similarity 83.3%; Pred. No. 18;
  Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WPWWPW 9
   | ||||
Db 6 WRWWPW 11

RESULT 6
T36208
hypothetical protein SCE36.09 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36208
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z21601
A;Accession: T36208
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-114 <OLI>
A;Cross-references: EMBL:AL049763; PIDN:CAB42078.1; GSPDB:GN00070; SCOEDB:SCE36.09
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE36.09

  Query Match 56.4%; Score 48.5; DB 2; Length 114;
  Best Local Similarity 80.0%; Pred. No. 5.1;
  Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 3 RW-PWWPWRR 11
   || | |||||
Db 103 RWRPRWPWRR 112

RESULT 7
S23449
NADH oxidase (H2O2-forming) (EC 1.6.-.-) - Thermus aquaticus
C;Species: Thermus aquaticus
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 23-Mar-1993
C;Accession: S23449; S24556
R;Park, H.J.; Kreutzer, R.; Reiser, C.O.A.; Sprinzl, M.
Eur. J. Biochem. 205, 875-879, 1992
A;Title: Molecular cloning and nucleotide sequence of the gene encoding a H(2)O(2)-formi
A;Reference number: S23449; MUID:92249331
A;Accession: S23449
A;Molecule type: DNA
A;Residues: 1-248 <PAR>
A;Cross-references: EMBL:X60110
A;Accession: S24556
A;Molecule type: protein
A;Residues: 1-32 <PAR1>
C;Genetics:
A;Gene: nox
C;Keywords: NAD; oxidoreductase
F;1-248/Product: NADH oxidase (H2O2-forming) #status experimental <MAT>

  Query Match 54.7%; Score 47; DB 2; Length 248;
  Best Local Similarity 100.0%; Pred. No. 17;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PWWPW 9
   |||||
Db 179 PWWPW 183

RESULT 8
G70715
hypothetical protein Rv0945 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70715
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: G70715
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-253 <COL>
A;Cross-references: GB:Z79700; GB:AL123456; NID:g3261628; PIDN:CAB02005.1; PID:g15242
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0945
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F;8-190/Domain: short-chain alcohol dehydrogenase homology <SADH>

  Query Match 54.7%; Score 47; DB 2; Length 253;
  Best Local Similarity 100.0%; Pred. No. 17;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PWWPW 9
   |||||
Db 230 PWWPW 234

RESULT 9
B83161
probable short-chain dehydrogenase PA3883 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83161
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: B83161
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <STO>

```
A;Cross-references: GB:AE004805; GB:AE004091; NID:g9950055; PIDN:AAG07270.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3883
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

  Query Match             54.7%;  Score 47;  DB 2;  Length 276;
  Best Local Similarity   70.0%;  Pred. No. 18;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 RWPWWPWRRK 12
           | |||| ||:
Db     197 RSPWWPLRRQ 206


RESULT  10
B70741
probable moeY protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: B70741
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70741
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-715 <COL>
A;Cross-references: GB:Z75555; GB:AL123456; NID:g3261608; PIDN:CAA99988.1; PID:e250356;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: moeY

  Query Match             54.7%;  Score 47;  DB 2;  Length 715;
  Best Local Similarity   66.7%;  Pred. No. 46;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 RWPWWPWRR 11
           || ::||||
Db      65 RWAYYPWRR 73


RESULT  11
T48529
hypothetical protein T22P22.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48529
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24490
A;Accession: T48529
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1411 <BEV>
A;Cross-references: EMBL:AL163814
A;Experimental source: cultivar Columbia; BAC clone T22P22
C;Genetics:
A;Map position: 5
A;Introns: 281/2; 320/1; 389/3; 429/3; 473/3; 515/3; 534/2; 567/3; 602/1; 669/1; 776/2;
A;Note: T22P22.90

  Query Match             54.7%;  Score 47;  DB 2;  Length 1411;
  Best Local Similarity   63.6%;  Pred. No. 87;
  Matches    7;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LRWPWWPWRRK 12
           | | |  ||||
Db    1013 LAWSWQQWRRK 1023


RESULT  12
T51071
related to trfA protein [imported] - Neurospora crassa
N;Alternate names: protein B2A19.50
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51071
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51071
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-728 <SCH>
A;Cross-references: EMBL:AL390092; GSPDB:GN00116; NCSP:B2A19.50
A;Experimental source: BAC clone B2A19; strain OR74A
C;Genetics:
A;Gene: NCSP:B2A19.50
A;Map position: 6
A;Introns: 26/1; 119/2

  Query Match             53.5%;  Score 46;  DB 2;  Length 728;
  Best Local Similarity   58.3%;  Pred. No. 62;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ILRWPWWPWRRK 12
           || |||  | |:
Db      11 ILGWPWLWWSRR 22


RESULT  13
A39128
anthranilate synthase (EC 4.1.3.27) component I [validated] - Pseudomonas syringae pv
N;Alternate names: anthranilate synthase alpha chain
C;Species: Pseudomonas syringae pv. savastanoi
C;Date: 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change 17-Mar-2000
C;Accession: A39128
R;da Costa, E.; Silva, O.; Kosuge, T.
J. Bacteriol. 173, 463-471, 1991
A;Title: Molecular characterization and expression analysis of the anthranilate synth
A;Reference number: A39128; MUID:91100331
A;Accession: A39128
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <DAC>
A;Cross-references: GB:M55911
C;Genetics:
A;Gene: trpE
C;Complex: heterotetramer; two component I chains, two component II chains
C;Function: <ANT>
A;Description: EC 4.1.3.27 [validated; MUID:90130325]
A;Pathway: tryptophan biosynthesis
A;Note: first step
C;Function: <COM1>
A;Description: EC 4.1.3.27 [validated; MUID:91100331]
A;Note: expression of .trpE seems to be independent of the concentration of tryptophan
C;Superfamily: anthranilate synthase component I
C;Keywords: carbon-carbon lyase; oxo-acid-lyase; tryptophan biosynthesis

  Query Match             52.9%;  Score 45.5;  DB 2;  Length 505;
  Best Local Similarity   28.0%;  Pred. No. 51;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels   13;  Gaps    1;

Qy       1 ILRW----------PWWPWRRK 12
             :| |           |||| |:
Db     467 VLEWEETLNKRRAMVGSAWWPWPRR 491
```

```
RESULT 14
S55483
modulator of drug activity homolog - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 10-Dec-1999
C;Accession: S55483; T38969
R;Connor, R.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S55479
A;Accession: S55483
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <CON>
A;Cross-references: EMBL:Z49811; NID:g854599; PIDN:CAA89955.1; PID:g854604
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, May 1995
A;Reference number: Z21821
A;Accession: T38969
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-196 <CO2>
A;Cross-references: EMBL:Z49811; PIDN:CAA89955.1; GSPDB:GN00066; SPDB:SPAC5H10.05c
A;Experimental source: strain 972h-; cosmid c5H10
C;Genetics:
A;Gene: SPDB:SPAC5H10.05c
A;Map position: 1
C;Superfamily: NAD(P)H dehydrogenase (quinone) 2

  Query Match             52.3%;  Score 45;  DB 2;  Length 196;
  Best Local Similarity   50.0%;  Pred. No. 24;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    4;  Gaps    2;

Qy       1 ILRWP-WW---PWRRK 12
           | :|| ||   ||: |
Db      63 IYQWPGWWMGTPWKLK 78


RESULT 15
F82646
monofunctional biosynthetic peptidoglycan transglycosylase XF1715 [imported] - Xylella f
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: F82646
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequer
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82646
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <SIM>
A;Cross-references: GB:AE003995; GB:AE003849; NID:g9106775; PIDN:AAF84524.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laign
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1715

  Query Match             52.3%;  Score 45;  DB 2;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWPWW 7
             |||||
Db      44 RWPWW 48


Search completed: January  4, 2002, 08:41:33
Job time: 171 sec
```

```
OM of: US-09-444-281-36 to: Issued_Patents_NA:*   out_format : pfs

Date: Jan 4, 2002 11:03 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084144_16216/app_query.fasta_1.210
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09444281_@CGN1_1_70 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-36
Query length: 12
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 135.410000



score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-071-49 +    75.00  169.84  0.0825     88  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-230-180-29 +    70.00  166.19  0.1319     39  !  S
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-259-741-5 +    70.00  125.59  24.07     6446  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-037-751-5 +    70.00  125.59  24.07     6446  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-466-422-5 +    70.00  125.59  24.07     6446  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-793-035-6 -    61.00  124.37  28.12      758  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-020-956-82 +    59.00  125.75  23.58      383  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-82 +    59.00  125.75  23.58      383  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-183 +    59.00  125.73  23.64      384  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-020-956-73 +    59.00  123.65  30.87      499  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-73 +    59.00  123.65  30.87      499  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-3 +    59.00  121.15  42.52      683  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-1 +    59.00  117.30  69.69     1109  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-6 +    59.00  117.30  69.69     1109  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-337A-1 +    58.00  112.83  123.69     1509  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-696-268B-1 +    58.00  112.83  123.69     1509  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05467-1 +    58.00  112.83  123.69     1509
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-11328A-3 +    58.00  112.38  130.96     1596
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-382-256-1 +    58.00  110.65  163.49     1984  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-395-115-1 +    58.00  110.65  163.49     1984  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-153-804-10 +    55.50  106.92  263.69     1677  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-159-784-1 -    55.00   98.94  734.12     4031  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-816-1 +    54.50  103.93  387.05     1894  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-179-147 +    54.00  114.83  95.61      423  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-915-246-1 +    54.00  104.75  348.65     1505  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-425-383-1 -    54.00   99.42  690.24     2941  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-920-812-13 +    54.00   90.09  2.3e+03     9515  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-920-827-13 +    54.00   90.09  2.3e+03     9515  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-921-177-13 +    54.00   90.09  2.3e+03     9515  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-362-577C-13 +    54.00   90.09  2.3e+03     9515  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-920-828-13 +    54.00   90.09  2.3e+03     9515  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2 +    54.00   41.31  5.7e+05     4403765
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-276-531-107 -    53.50   98.10  817.85     3058  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-312-038-4 -    53.00  103.92  387.54     1294  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-312-038-3 -    53.00   99.39  693.17     2289  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-417-1 +    53.00   91.29  2.0e+03     6344  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-057-860A-1 +    52.50  108.90  204.66      609  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-026-958-1 +    52.50  100.95  567.45     1656  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-808-931-14 -    52.50   94.25  1.3e+03     3848  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-808-323-14 -    52.50   94.25  1.3e+03     3848  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-050-603A-14 -    52.50   94.25  1.3e+03     3848
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-420B-14 -    52.50   94.25  1.3e+03     3848
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-149-097D-24 +    52.50   89.45  2.5e+03     7032
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-949-386-24 +    52.50   89.45  2.5e+03     7032
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-450-562-24 +    52.50   89.45  2.5e+03     7032

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-042-071-49

seq_documentation_block:
; Sequence 49, Application US/09042071
; Patent No. 6294372
;  GENERAL INFORMATION:
;    APPLICANT:  Burian, Jan
;    APPLICANT:  Kay, William W.
;    TITLE OF INVENTION:  REPLICATION GENES AND GENE PRODUCTS FROM
;    TITLE OF INVENTION:  SMALL CRYPTIC PLASMIDS AND METHODS FOR CONSTRUCTING
;    TITLE OF INVENTION:  CONTROLLED-REPLICATION PLASMID VECTORS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,071
;      FILING DATE:  13-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  660081.407
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  88 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-042-071-49


alignment_scores:
            Quality:   75.00              Length:      10
              Ratio:  7.500                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
 US-09-444-281-36 x US-09-042-071-49   ..

 Align seg 1/1  to: US-09-042-071-49  from: 1  to: 88

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         :::||||||||||||||||||||||||||||
      34 AAATGGCCGTGGTGGCCGTGGCGTCGTAAA 63

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-230-180-29

seq_documentation_block:
; Sequence 29, Application US/09230180
; Patent No. 6183992
; GENERAL INFORMATION:
;  APPLICANT:  Kim, Sun-Chang
;  APPLICANT:  Lee, Jae Hyun
;  APPLICANT:  Kang, Min Hyung
```

```
;  APPLICANT:  Kim, Jeong Hyun
;  APPLICANT:  Hong, Seung-Suh
;  APPLICANT:  Lee, Hyun-Soo
;  APPLICANT:  Samyang Genex Corporation
;  APPLICANT:  Korea Advanced Institute of Science and Technology
;  TITLE OF INVENTION:  METHOD FOR MASS PRODUCTION OF
;  TITLE OF INVENTION:  ANTIMICROBIAL PEPTIDE
;  FILE REFERENCE:  6181/0F135
;  CURRENT APPLICATION NUMBER: US/09/230,180
;  CURRENT FILING DATE:  1999-03-10
;  PRIOR APPLICATION NUMBER:  PCT/KR98/00132
;  PRIOR FILING DATE:  1998-05-28
;  PRIOR APPLICATION NUMBER:  KR 13372/1998
;  PRIOR FILING DATE:  1998-04-09
;  PRIOR APPLICATION NUMBER:  KR 21312/1997
;  PRIOR FILING DATE:  1997-05-28
;  NUMBER OF SEQ ID NOS:  36
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 39
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: DNA sequence deduced from Indolicidin peptide
;   OTHER INFORMATION: sequence based on codon usage of E. coli
US-09-230-180-29


alignment_scores:
            Quality:   70.00              Length:      9
              Ratio:  7.778                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x US-09-230-180-29   ..

 Align seg 1/1  to: US-09-230-180-29  from: 1  to: 39

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
      13 AAATGGCCGTGGTGGCCGTGGCGTCGT 39

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-259-741-5

seq_documentation_block:
; Sequence 5, Application US/09259741
; Patent No. 6033895
;  GENERAL INFORMATION:
;    APPLICANT:  GARGER, STEPHEN
;    APPLICANT:  HOLTZ, R. BARRY
;    APPLICANT:  McCULLOCH, MICHAEL
;    APPLICANT:  TURPEN, THOMAS
;    TITLE OF INVENTION:  A PROCESS FOR ISOLATING AND
;    TITLE OF INVENTION:  PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
;    TITLE OF INVENTION:  SOURCES
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howrey & Simon
;      STREET:  1299 Pennsylvania Avenue N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/259,741
;      FILING DATE:  February 25, 1999
;      CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/037,751
;      FILING DATE:  March 10, 1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P
;      REGISTRATION NUMBER:  25,277
;      REFERENCE/DOCKET NUMBER:  00801.0140.US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-463-8100
;      TELEFAX:  650-463-8400
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6446 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  Genomic RNA
US-09-259-741-5


alignment_scores:
            Quality:   70.00              Length:      9
              Ratio:  7.778                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x US-09-259-741-5   ..

 Align seg 1/1  to: US-09-259-741-5  from: 1  to: 6446

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-037-751-5

seq_documentation_block:
; Sequence 5, Application US/09037751
; Patent No. 6037456
;  GENERAL INFORMATION:
;    APPLICANT:  GARGER, STEPHEN
;    APPLICANT:  HOLTZ, R. BARRY
;    APPLICANT:  McCULLOCH, MICHAEL
;    APPLICANT:  TURPEN, THOMAS
;    TITLE OF INVENTION:  A PROCESS FOR ISOLATING AND
;    TITLE OF INVENTION:  PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
;    TITLE OF INVENTION:  FROM PLANT SOURCES
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howrey & Simon
;      STREET:  1299 Pennsylvania Avenue N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,751
;      FILING DATE:  10-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P
;      REGISTRATION NUMBER:  25,277
;      REFERENCE/DOCKET NUMBER:  00801.0140.999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-463-8109
```

```
;       TELEFAX:  650-463-8400
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6446 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  Genomic RNA
US-09-037-751-5


alignment_scores:
            Quality:   70.00          Length:      9
              Ratio:  7.778            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x US-09-037-751-5    ..

 Align seg 1/1  to: US-09-037-751-5  from: 1  to: 6446

      3 ArgTrpProTrpTrpProTrpArgArg 11
        :::|||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-466-422-5

seq_documentation_block:
; Sequence 5, Application US/09466422
; Patent No. 6303779
;   GENERAL INFORMATION:
;       APPLICANT: GARGER, STEPHEN
;                  HOLTZ, R. BARRY
;                  McCULLOCH, MICHAEL
;                  TURPEN, THOMAS
;       TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
;                           PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
;                           FROM PLANT SOURCES
;       NUMBER OF SEQUENCES: 5
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howrey & Simon
;            STREET: 1299 Pennsylvania Avenue N.W.
;            CITY: Washington
;            STATE: DC
;            COUNTRY: USA
;            ZIP: 20004
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/466,422
;            FILING DATE: 17-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/037,751
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Halluin, Albert P
;            REGISTRATION NUMBER: 25,277
;            REFERENCE/DOCKET NUMBER: 00801.0140.999
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650-463-8109
;            TELEFAX: 650-463-8400
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO:  5:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 6446 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: unknown
;       MOLECULE TYPE: Genomic RNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-466-422-5


alignment_scores:
            Quality:   70.00          Length:      9
              Ratio:  7.778            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x US-09-466-422-5    ..

 Align seg 1/1  to: US-09-466-422-5  from: 1  to: 6446

      3 ArgTrpProTrpTrpProTrpArgArg 11
        :::|||||||||||||||||||||||||
    6213 AAGUGGCCUUGGUGGCCAUGGCGCCGA 6239

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-793-035-6

seq_documentation_block:
; Sequence 6, Application US/08793035
; Patent No. 6011201
;  GENERAL INFORMATION:
;     APPLICANT:  Slabas, Antoni R.
;     APPLICANT:  White, Andrew
;     APPLICANT:  Chase, Dianne
;     APPLICANT:  Elborough, Keiran
;     APPLICANT:  Fentem, Phillip A.
;     TITLE OF INVENTION:  B-Ketoacyl ACP Reductase Genes From
;     TITLE OF INVENTION:  Brassica Napus
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  TX
;       COUNTRY:  US
;       ZIP:  77210-4433
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/793,035
;       FILING DATE:  28-JUL-1997
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9414622.2
;       FILING DATE:  20-JUL-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB95/01678
;       FILING DATE:  17-JUL-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kammerer, Patricia A.
;       REGISTRATION NUMBER:  29,775
;       REFERENCE/DOCKET NUMBER:  MOBT:132
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713.787.1400
;       TELEFAX:  713.787.1440
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  758 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-793-035-6


alignment_scores:
            Quality:   61.00          Length:      9
```

```
            Ratio:  7.625                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-444-281-36 x US-08-793-035-6/rev   ..

 Align seg 1/1  to reverse of: US-08-793-035-6  from: 1  to: 758

       1 IleLeuArgTrpProTrpTrpProTrp 9
         :::|||||||||   ||||||||||
     451 GTCCTGCGATGGAGGTGGTGGCCATGG 425

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-020-956-82

seq_documentation_block:
; Sequence 82, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  82:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  383 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-82


alignment_scores:
            Quality:   59.00              Length:      8
              Ratio:  8.429                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 87.500

alignment_block:
 US-09-444-281-36 x US-09-020-956-82   ..

 Align seg 1/1  to: US-09-020-956-82  from: 1  to: 383

       2 LeuArgTrpProTrpTrpProTrp 9
         ||||||||||||||||||   |||
     155 CTTCGCTGGCCTTGGTGGAGCTGG 178

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-82

seq_documentation_block:
; Sequence 82, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  82:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  383 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-82


alignment_scores:
            Quality:   59.00              Length:      8
              Ratio:  8.429                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 87.500

alignment_block:
 US-09-444-281-36 x US-09-030-607-82   ..

 Align seg 1/1  to: US-09-030-607-82  from: 1  to: 383

       2 LeuArgTrpProTrpTrpProTrp 9
         ||||||||||||||||||   |||
     155 CTTCGCTGGCCTTGGTGGAGCTGG 178

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-183

seq_documentation_block:
; Sequence 183, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
```

```
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/030,607
;       FILING DATE:  25-FEB-1998
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  210121.427C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  183:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  384 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-030-607-183


alignment_scores:
            Quality:   59.00              Length:      8
              Ratio:  8.429                Gaps:       0
 Percent Similarity:  87.500   Percent Identity: 87.500

alignment_block:
 US-09-444-281-36 x US-09-030-607-183    ..

 Align seg 1/1  to: US-09-030-607-183  from: 1  to: 384

       2 LeuArgTrpProTrpTrpProTrp 9
         ||||||||||||||||||    |||
     156 CTTCGCTGGCCTTGGTGGAGCTGG 179

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-020-956-73

seq_documentation_block:
; Sequence 73, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  73:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-73


alignment_scores:
            Quality:   59.00              Length:      8
              Ratio:  8.429                Gaps:       0
 Percent Similarity:  87.500   Percent Identity: 87.500

alignment_block:
 US-09-444-281-36 x US-09-020-956-73    ..

 Align seg 1/1  to: US-09-020-956-73  from: 1  to: 499

       2 LeuArgTrpProTrpTrpProTrp 9
         ||||||||||||||||||    |||
     115 CTTCGCTGGCCTTGGTGGAGCTGG 138

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-030-607-73

seq_documentation_block:
; Sequence 73, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  73:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-73


alignment_scores:
            Quality:   59.00              Length:      8
```

```
            Ratio:  8.429              Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 87.500

alignment_block:
 US-09-444-281-36 x US-09-030-607-73   ..

 Align seg 1/1  to: US-09-030-607-73  from: 1  to: 499

       2 LeuArgTrpProTrpTrpProTrp 9
         ||||||||||||||||||   |||
     115 CTTCGCTGGCCTTGGTGGAGCTGG 138

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-3

seq_documentation_block:
; Sequence 3, Application US/09088651
; Patent No. 6165771
;  GENERAL INFORMATION:
;    APPLICANT:  BURGESS, NICOLA A.
;    APPLICANT:  CLINKENBEARD, HELEN E.
;    APPLICANT:  SOUTHAN, CHRISTOPHER D.
;    TITLE OF INVENTION:  NOVEL COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/088,651
;      FILING DATE:  JUNE 1, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB9712088.5
;      FILING DATE:  10-JUNE-1997
;      APPLICATION NUMBER:  EP 97308295.1
;      FILING DATE:  17-OCT-1997
;      APPLICATION NUMBER:  GB 9803650.2
;      FILING DATE:  20-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F.
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH30358
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  683 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-088-651-3


alignment_scores:
            Quality:   59.00              Length:      11
              Ratio:  6.556                Gaps:       0
 Percent Similarity:  81.818   Percent Identity: 63.636

alignment_block:
 US-09-444-281-36 x US-09-088-651-3   ..

 Align seg 1/1  to: US-09-088-651-3  from: 1  to: 683

       2 LeuArgTrpProTrpTrpProTrpArgArgLys 12
         |||   ||||||||||||   ||||||::::::
     139 CTGACGTGGCCCTGGTGGAGCTGGAGGCACCAG 171

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-1

seq_documentation_block:
; Sequence 1, Application US/09088651
; Patent No. 6165771
;  GENERAL INFORMATION:
;    APPLICANT:  BURGESS, NICOLA A.
;    APPLICANT:  CLINKENBEARD, HELEN E.
;    APPLICANT:  SOUTHAN, CHRISTOPHER D.
;    TITLE OF INVENTION:  NOVEL COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/088,651
;      FILING DATE:  JUNE 1, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB9712088.5
;      FILING DATE:  10-JUNE-1997
;      APPLICATION NUMBER:  EP 97308295.1
;      FILING DATE:  17-OCT-1997
;      APPLICATION NUMBER:  GB 9803650.2
;      FILING DATE:  20-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F.
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH30358
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1109 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-088-651-1


alignment_scores:
            Quality:   59.00              Length:      11
              Ratio:  6.556                Gaps:       0
 Percent Similarity:  81.818   Percent Identity: 63.636

alignment_block:
 US-09-444-281-36 x US-09-088-651-1   ..

 Align seg 1/1  to: US-09-088-651-1  from: 1  to: 1109

       2 LeuArgTrpProTrpTrpProTrpArgArgLys 12
         |||   ||||||||||||   ||||||::::::
     564 CTTACGTGGCCCTGGTGGAGCTGGAGGCACCAG 596

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-088-651-6
```

```
seq_documentation_block:
; Sequence 6, Application US/09088651
; Patent No. 6165771
;  GENERAL INFORMATION:
;    APPLICANT:  BURGESS, NICOLA A.
;    APPLICANT:  CLINKENBEARD, HELEN E.
;    APPLICANT:  SOUTHAN, CHRISTOPHER D.
;    TITLE OF INVENTION:  NOVEL COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/088,651
;      FILING DATE:  JUNE 1, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB9712088.5
;      FILING DATE:  10-JUNE-1997
;      APPLICATION NUMBER:  EP 97308295.1
;      FILING DATE:  17-OCT-1997
;      APPLICATION NUMBER:  GB 9803650.2
;      FILING DATE:  20-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F.
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH30358
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1109 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-088-651-6


alignment_scores:
            Quality:   59.00              Length:     11
              Ratio:  6.556                Gaps:      0
 Percent Similarity:  81.818   Percent Identity: 63.636

alignment_block:
 US-09-444-281-36 x US-09-088-651-6  ..

 Align seg 1/1  to: US-09-088-651-6  from: 1  to: 1109

       2 LeuArgTrpProTrpTrpProTrpArgArgLys 12
         |||   ||||||||||||   ||||||::::::
     565 CTGACGTGGCCCTGGTGGAGCTGGAGGCACCAG 597

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-481-337A-1

seq_documentation_block:
; Sequence 1, Application US/08481337A
; Patent No. 5863738
;  GENERAL INFORMATION:
;    APPLICANT:  TEN DIJKE, Peter
;    APPLICANT:  HELDIN, Carl-Henrik
;    APPLICANT:  MIYAZONO, Kohei
;    APPLICANT:  SAMPATH, Kuber T.
;    TITLE OF INVENTION:  Morphogenic Protein-Specific Cell
;    TITLE OF INVENTION:  Surface Receptors and Uses Therefor
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,337A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  CRP-097CP2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1509
;      OTHER INFORMATION:  /product= "Human ALK1"
US-08-481-337A-1


alignment_scores:
            Quality:   58.00              Length:      6
              Ratio:  9.667                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-444-281-36 x US-08-481-337A-1  ..

 Align seg 1/1  to: US-08-481-337A-1  from: 1  to: 1509

       4 TrpProTrpTrpProTrp 9
         ||||||||||||||||||
     389 TGGCCCTGGTGGCCCTGG 406
```

```
OM of: US-09-444-281-36 to: GenEmbl:* , out_format : pfs

Date: Jan 4, 2002 10:56 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09444281/runat_04012002_084143_16195/app_query.fasta_1.210
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09444281_@CGN1_1_5145 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-444-281-36
Query length: 12
Database: GenEmbl:*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1728.450000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_om:BTINDLCD	+	70.00	149.45	4.77	550	! X67340 B.taurus mRNA for indoli
gb_pat:AX098418	+	70.00	132.44	42.28	6446	! AX098418 Sequence 5 from Paten
gb_ro:MMU08210	+	69.00	136.50	25.10	2651	! U08210 Mus musculus tropoelast
gb_ro:AF289665	-	69.00	110.92	667.65	107257	! AF289665 Mus musculus EIF4H
gb_htg:AP003754	-	69.00	109.64	786.64	129052	! AP003754 Oryza sativa chromc
gb_htg:AC091250	+	69.00	106.59	1.2e+03	200849	! AC091250 Mus musculus chromc
gb_ro:AF210429	+	67.00	144.52	8.98	456	! AF210429 Mus musculus group X s
gb_ro:AF166097	+	67.00	138.82	18.65	1040	! AF166097 Mus musculus group X
gb_pl:ATPHYB	+	64.50	124.58	115.87	3850	! X17342 Arabidopsis thaliana pH
gb_pl:ATHPHTOCHB	+	64.50	120.95	184.57	6509	! L09262 Arabidopsis thaliana pH
gb_pl:AC005724	+	64.50	103.05	1.8e+03	86671	! AC005724 Arabidopsis thaliana
gb_pl:AP003141	-	64.50	99.67	2.8e+03	141275	! AP003141 Oryza sativa genomi
gb_ro:MMA2IXCOA	-	64.00	112.33	557.28	19479	! Z22923 M.musculus alpha2 (IX)
gb_htg:AL356097_0	-	64.00	100.36	2.6e+03	110000	! AL356097 Homo sapiens chromc
gb_pl:AP001366	+	64.00	98.40	3.3e+03	146081	! AP001366 Oryza sativa genomi
gb_htg:AC027113	+	64.00	97.80	3.6e+03	159391	! AC027113 Homo sapiens chromc
gb_htg:AL358473	+	64.00	96.66	4.2e+03	187845	! AL358473 Homo sapiens chromc
gb_ro:AF191309	-	63.00	122.71	147.17	3213	! AF191309 Mus musculus zinc fin
gb_ba:RRPHAS	-	63.00	121.42	173.77	3875	! X66407 R.ruber ORF1, phbCRr, C
gb_ro:RNY17153	+	63.00	118.97	237.92	5523	! Y17153 Rattus norvegicus mRNA
gb_ro:RATNCHVS	+	63.00	117.37	291.95	6957	! M26643 Rat skeletal muscle vol
gb_in:CEF46A8	+	63.00	111.19	645.18	17015	! Z81539 Caenorhabditis elegans
gb_ba:AE004649	+	63.00	108.98	856.61	23425	! AE004649 Pseudomonas aerugino
gb_htg:AC020365	+	63.00	107.80	996.42	27780	! AC020365 Drosophila melanogas
gb_in:AC008316	+	63.00	95.66	4.7e+03	160817	! AC008316 Drosophila melanoga
gb_in:AC008315	+	63.00	95.00	5.1e+03	177028	! AC008315 Drosophila melanoga
gb_in:AE003684	+	63.00	93.51	6.2e+03	219579	! AE003684 Drosophila melanoga
gb_htg:AC068379	+	63.00	93.41	6.3e+03	222876	! AC068379 Homo sapiens chromc
gb_htg:AC091358	+	62.00	94.51	5.5e+03	140738	! AC091358 Rattus norvegicus c
gb_pl:FSGRPF1	+	61.50	130.96	51.10	621	! X98539 F.sylvatica mRNA for GRF
gb_ba:AE005916	+	61.50	110.64	692.58	11745	! AE005916 Caulobacter crescent
gb_pl:AP002872	-	61.50	92.71	6.9e+03	157112	! AP002872 Oryza sativa genomi
gb_pat:A48353	-	61.00	128.53	69.75	759	! A48353 Sequence 6 from Patent W
gb_pl:BNA243084	-	61.00	124.95	110.41	1274	! AJ243084 Brassica napus mRNA f
gb_pr:AF112471	-	61.00	122.76	146.30	1750	! AF112471 Homo sapiens calcium/
gb_pr:AF112472	-	61.00	122.47	151.84	1825	! AF112472 Homo sapiens calcium/
gb_pr:D50604	+	61.00	121.82	165.12	2006	! D50604 Human beta-cytoplasmic
gb_ba:ASAJ187	-	61.00	114.00	449.90	6213	! AJ000187 Arthrobacter sp. catA
gb_ba:AF319543	-	61.00	113.25	495.45	6927	! AF319543 Streptomyces venezuel
gb_htg:AC013145	+	61.00	103.19	1.8e+03	29692	! AC013145 Drosophila melanogas
gb_pr:AL158051	-	61.00	98.37	3.3e+03	59585	! AL158051 Human DNA sequence f
gb_pl:AB046440	+	61.00	95.03	5.1e+03	96684	! AB046440 Arabidopsis thali
gb_in:DMBR37P7	-	61.00	94.61	5.4e+03	102619	! AL050231 Drosophila melan
gb_pl:AC051625	-	61.00	93.74	6.0e+03	116461	! AC051625 Genomic Sequence
gb_htg:AC010475	-	61.00	93.25	6.4e+03	124927	! AC010475 Homo sapiens chr

```
seq_name: gb_om:BTINDLCD

seq_documentation_block:
LOCUS       BTINDLCD      550 bp    mRNA            MAM       07-OCT-1992
DEFINITION  B.taurus mRNA for indolicidin.
ACCESSION   X67340
VERSION     X67340.1  GI:462
KEYWORDS    indolicidin.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Del Sal,G., Storici,P., Schneider,C., Romeo,D. and Zanetti,M.
  TITLE     cDNA cloning of the neutrophil bactericidal peptide indolicidin
  JOURNAL   Biochem. Biophys. Res. Commun. 187 (1), 467-472 (1992)
  MEDLINE   92392368
REFERENCE   2  (bases 1 to 550)
  AUTHORS   Del Sal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1992) G. Del Sal, Univ. of Trieste, Dip. di
            Biochimica, Biofisica e Chimica, delle Macromolecole, Via A.
            Valerio, 38, 34127 Trieste, ITALY
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /cell_line="bone marrow cells"
     sig_peptide     13. .99
     CDS             13. .447
                     /codon_start=1
                     /product="indolicin prepropeptide"
                     /protein_id="CAA47755.1"
                     /db_xref="GI:463"
                     /db_xref="SWISS-PROT:P33046"
                     /translation="MQTQRASLSLGRWSLWLLLLGLVVPSASAQALSYREAVLRAVDQ
                     LNELSSEANLYRLLELDPPPKDNEDLGTRKPVSFTVKETVCPRTIQQPAEQCDFKEKG
                     RVKQCVGTVTLDPSNDQFDLNCNELQSVILPWKWPWWPWRRG"
     mat_peptide     403. .444
                     /product="indolicin"
BASE COUNT      126 a    140 c    169 g    115 t
ORIGIN


alignment_scores:
            Quality:   70.00            Length:      9
              Ratio:  7.778              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x BTINDLCD   ..

 Align seg 1/1  to: BTINDLCD  from: 1  to: 550

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::||||||||||||||||||||||||
     415 AAATGGCCATGGTGGCCTTGGCGCAGA 441

seq_name: gb_pat:AX098418

seq_documentation_block:
LOCUS       AX098418     6446 bp    mRNA            PAT       03-APR-2001
DEFINITION  Sequence 5 from Patent WO0119969.
ACCESSION   AX098418
VERSION     AX098418.1  GI:13537710
KEYWORDS    .
SOURCE      Nicotiana benthamiana.
```

```
  ORGANISM  Nicotiana benthamiana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (bases 1 to 6446)
  AUTHORS   Garger,S.J., Holtz,B.R., Mcculloch,M.J. and Turpen,T.H.
  TITLE     A process for isolating and purifying viruses, soluble proteins and
            peptides from plant sources
  JOURNAL   Patent: WO 0119969-A 5 22-MAR-2001;
            Large Scale Biology Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .6446
                     /organism="Nicotiana benthamiana"
                     /db_xref="taxon:4100"
BASE COUNT     1873 a   1234 c   1563 g   1776 t
ORIGIN


alignment_scores:
            Quality:   70.00            Length:      9
              Ratio:  7.778              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
 US-09-444-281-36 x AX098418   ..

 Align seg 1/1  to: AX098418  from: 1  to: 6446

       3 ArgTrpProTrpTrpProTrpArgArg 11
         :::|||||||||||||||||||||||||
    6213 AAGTGGCCTTGGTGGCCATGGCGCCGA 6239

seq_name: gb_ro:MMU08210

seq_documentation_block:
LOCUS       MMU08210     2651 bp    mRNA            ROD       31-OCT-1995
DEFINITION  Mus musculus tropoelastin mRNA, complete cds.
ACCESSION   U08210
VERSION     U08210.1  GI:473273
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2651)
  AUTHORS   Wydner,K.S., Sechler,J.L., Boyd,C.D. and Passmore,H.C.
  TITLE     Use of an intron polymorphism to localize the tropoelastin gene to
            mouse chromosome 5 in a region of linkage conservation with human
            chromosome 7
  JOURNAL   Genomics 23 (1), 125-131 (1994)
  MEDLINE   95130069
REFERENCE   2  (bases 1 to 2651)
  AUTHORS   Boyd,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1994) Charles D. Boyd, Department of Surgery,
            UMDNJ - Robert Wood Johnson Medical School, 51 French St., New
            Brunswick, NJ 08903, USA
FEATURES             Location/Qualifiers
     source          1. .2651
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="lung"
                     /dev_stage="adult"
     CDS             1. .2583
                     /codon_start=1
                     /product="tropoelastin"
                     /protein_id="AAA80155.1"
                     /db_xref="GI:473274"
                     /translation="MAGLTAVVPQPGVLLILLLNLLHPAQPGGVPGAVPGGLPGGVPG
                     GVYYPGAGGGLGGGGGALGPGGKPPKPGAGLLGTFGAGPGGLGGAGPGAGLGAFPAG
                     TFPGAGAVPGGAAGAAAAYKAAAKAGAGLGGVGGVPGGVGVGGVPGGVGVGGVPGGV
                     GVGGVPGGVGGIGGIGGLGVSTGAVVPQVGAGIGAGGKPGKVPGVGLPGVYPGGVLPG
                     TGARFPGVGVLPGVPTGTGVKAKAPGGGGAFSGIPGVGPFGGQQPGVPLGYPIKAPKL
                     PGGYGLPYTNGKLPYGVAGAGGKAGYPTGTGVGSQAAAAAAKAAKYGAGGAGVLPGVG
                     GGGIPGGAGAIPGIGGIAGAGTPAAAAAAKAAAKAAKYGAAGGLVPGGPGVRLPGAGI
                     PGVGGIPGVGGIPGVGGPGIGGPGIVGGPGAVSPAAAAKAAAKAAKYGARGGVGIPTY
                     GVGAGGFPGYGVGAGAGLGGASPAAAAAAAKAAKYGAGGAGALGGLVPGAVPGALPGA
                     VPAVPGAGGVPGAGTPAAAAAAAAAKAAAKAGLGPGVGGVPGGVGVGGIPGGVGVGGV
                     PGGVGPGGVTGIGAGPGGLGGAGSPAAAKSAAKAAAKAQYRAAAGLGAGVPGFGAGAG
                     VPGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGGLGGPG
                     GLGGPGGLGGPGGLGGAGVPGRVAGAAPPAAAAAAAKAAAKAAQYGLGGAGGLGAGGL
                     GAGGLGAGGLGAGGLGAGGLGAGGLGAGGGVSPAAAAKAAKYGAAGLGGVLGA
                     RPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPYGGALGALGYQGGGCFGKSCG
                     RKRK"
BASE COUNT     410 a    623 c   1029 g    589 t
ORIGIN


alignment_scores:
            Quality:   69.00            Length:     10
              Ratio:  7.667              Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x MMU08210   ..

 Align seg 1/1  to: MMU08210  from: 1  to: 2651

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         |||||||||||||||||||||||||   :::
    2025 AGGTGGCCTTGGTGGCCCTGGAGGTCTCGG 2054

seq_name: gb_ro:AF289665

seq_documentation_block:
LOCUS       AF289665   107257 bp    DNA             ROD       14-AUG-2000
DEFINITION  Mus musculus EIF4H gene, partial cds; LIMK1 gene, complete cds; and
            ELN gene, partial cds.
ACCESSION   AF289665
VERSION     AF289665.1  GI:9800517
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 107257)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Genome Technology Branch, National Human
            Genome Research Institute, 49 Convent Dr. Rm. 2A02, Bethesda, MD
            20892, USA
FEATURES             Location/Qualifiers
     source          1. .107257
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="42j20"
                     /clone_lib="Research Genetics CITB-CJ7-B BAC library"
     mRNA            complement(<7676. .7763)
                     /product="EIF4H"
     CDS             complement(<7676. .7734)
                     /codon_start=1
                     /product="EIF4H"
                     /protein_id="AAF99335.1"
                     /db_xref="GI:9800519"
                     /translation="MADFDTYDDRAYSSFGGGRG"
     mRNA            complement(join(24480. .25812,25920. .26077,26336. .26391,
                     28274. .28430,30220. .30285,30530. .30589,32612. .32743,
                     33634. .33720,34257. .34440,37341. .37507,37617. .37722,
                     38745. .38951,41098. .41207,41439. .41577,55071. .55167,
                     56768. .57026))
                     /product="LIMK1"
     CDS             complement(join(25650. .25812,25920. .26077,26336. .26391,
```

```
                     28274. .28430,30220. .30285,30530. .30589,32612. .32743,
                     33634. .33720,34257. .34440,37341. .37507,37617. .37722,
                     38745. .38951,41098. .41207,41439. .41577,55071. .55167,
                     56768. .56822))
                     /codon_start=1
                     /evidence=not_experimental
                     /product="LIMK1"
                     /protein_id="AAF99334.1"
                     /db_xref="GI:9800518"
                     /translation="MRLTLLCCTWREERMGEEGSELPVCASCGQRIYDGQYLQALNAD
                     WHADCFRCCECSVSLSHQYYEKDGQLFCKKDYWARYGESCHGCSEHITKGLVMVAGEL
                     KYHPECFICLACGNFIGDGDTYTLVEHSKLYCGQCYYQTVVTPVIEQILPDSPGSHLP
                     HTVTLVSIPASAHGKRGLSVSIDPPHGPPGCGTEHSHTVRVQGVDPGCMSPDVKNSIH
                     VGDRILEINGTPIRNVPLDEIDLLIQETSRLLQLTLEHDPHDSLGHGPVSDPSPLSSP
                     VHTPSGQAASSARQKPVLRSCSIDTSPGTSSLASPASQRKDLGRSESLRVVCRPHRIF
                     RPSDLIHGEVLGKGCFGQAIKVTHRETGEVMVMKELIRFDEETQRTFLKEVKVMRCLE
                     HPNVLKFIGVLYKDKRLNFITEYIKGGTLRGIIKNMDSQYPWSQRVSFAKDIASGMAY
                     LHSMNIIHRDLNSHNCLVRENRNVVVADFGLARLMIDEKNQSEDLRSLKKPDRKKRYT
                     VVGNPYWMAPEMINGRSYDEKVDVFSFGIVLCEIIGRVNADPDYLPRTMDFGLNVRGF
                     LDRYCPPNCPPSFFPITVRCCDLDPEKRPSFVKLEQWLETLRMHLSGHLPLGPQLEQL
                     ERGFWETYRRGESSLPAHPEVPD"
mRNA                 complement(join(72144. .72255,72961. .73011,73719. .73748,
                     74172. .74216,74796. .74846,74997. .75035,75941. .76081,
                     76252. .76314,76403. .76504,78538. .78576,79493. .79651,
                     80837. .80896,80981. .81121,82426. .82485,83678. .83779,
                     84115. .84162,84859. .84948,85423. .85476,86452. .86598,
                     87420. .87479,87564. .87653,90048. .90095,90466. .90510,
                     90895. .90936,93350. .93421,94200. .94229,94899. .94970,
                     96821. .96868,97559. .97621,97702. .97863,98076. .98177,
                     99496. .99531,100964. .100993,104734. .104766,
                     105595. .>105630))
                     /note="GenScan; P1=0.902, P2=0.979, P3=0.899, P4=0.797,
                     P5=0.640, P6=0.421, P7=0.897, P8=0.912, P9=0.991,
                     P10=0.999, P11=0.630, P12=0.813, P13=0.935, P14=0.767,
                     P15=0.773, P16=0.567, P17=0.806, P18=0.998, P19=0.999,
                     P20=0.996, P21=0.978, P22=0.980, P23=0.999, P24=0.986,
                     P25=0.579, P26=0.757, P27=0.661, P28=0.985, P29=0.952,
                     P30=0.228, P31=0.203, P32=0.451, P33=0.463, P34=0.525,
                     P35=0.228, P36=0.969, P37=0.995, P38=0.905, P39=0.655,
                     P40=0.986, P41=0.973, P42=0.930, P43=0.967, P44=0.995,
                     P45=0.999, P46=0.351, P47=0.994, P48=0.534, P49=0.787,
                     P50=0.943, P51=PlyA"
                     /product="ELN"
CDS                  complement(join(72212. .72255,72961. .73011,73719. .73748,
                     74172. .74216,74796. .74846,74997. .75035,75941. .76081,
                     76252. .76314,76403. .76504,78538. .78576,79493. .79651,
                     80837. .80896,80981. .81121,82426. .82485,83678. .83779,
                     84115. .84162,84859. .84948,85423. .85476,86452. .86598,
                     87420. .87479,87564. .87653,90048. .90095,90466. .90510,
                     90895. .90936,93350. .93421,94200. .94229,94899. .94970,
                     96821. .96868,97559. .97621,97702. .97863,98076. .98177,
                     99496. .99531,100964. .100993,104734. .104766,
                     105595. .>105630))
                     /codon_start=3
                     /product="ELN"
                     /protein_id="AAF99336.1"
                     /db_xref="GI:9800520"
                     /translation="AGIGGLGGGGGALGPGGKPPKPGAGLLGTFGAGPGGLGGAGPGA
                     GLGAFPAGTFPGAGALVPGGAAGAAAAYKAAAKAGAGLGGVGGVPGGVGVGGVPGGVG
                     VGGVPGGVGVGGVPGGVGGIGGIGGLGVSTGAVVPQVGAGIGAGGKPGKVPGVGLPGV
                     YPGGVLPGTGARFPGVGVLPGVPTGTGVKAKAPGGGGAFAGIPGVGPFGGQQPGVPLG
                     YPIKAPKLPGGYGLPYTNGKLPYGVAGAGGKAGYPTGTGVGSQAAAAAAKAAKYGAGG
                     AGVLPGVGGGGIPGGAGAIPGIGGIAGAGTPAAAAAAKAAAKAAKYGAAGGLVPGGPG
                     VRLPGAGIPGVGGIPGVGGIPGVGGPGIGGPGIVGGPGAVSPAAAAKAAAKAAKYGAR
                     GGVGIPTYGVGAGGFPGYGVGAGAGLGGASPAAAAAAAKAAKYGAGGAGALGGLVPGA
                     VPGALPGAVPAVPGAGGVPGAGTPAAAAAAAAAKAAAKAGLGPGVGGVPGGVGVGGIP
                     GGVGVGGVPGGVGPGGVTGIGAGPGGLGGAGSPAAAKSAAKAAAKAQYRAAAGLGAGV
                     PGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGAVPGSLAASKAAKYGA
                     AGGLGGPGGLGGPGGLGGPGGLGGAGVPGRVAGAAPPAAAAAAAKAAAKAAQYGLGGA
                     GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGGVSPAAAAKAAKYGAA
                     GLGGVLGARPFPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPYGGALGALGYQGG
                     GCFGKSCGRKRK"
BASE COUNT       26300 a  26398 c  26598 g  27966 t
```

```
ORIGIN


alignment_scores:
            Quality:   69.00              Length:     10
              Ratio:  7.667                Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AF289665/rev   ..

 Align seg 1/1  to reverse of: AF289665  from: 1  to: 107257

       3 ArgTrpProTrpTrpProTrpArgArgLys 12
         ||||||||||||||||||||||||||   :::
   76497 AGGTGGCCTTGGTGGCCCTGGAGGTCTCGG 76468

seq_name: gb_htg:AP003754

seq_documentation_block:
LOCUS        AP003754   129052 bp    DNA              HTG       14-JUN-2001
DEFINITION   Oryza sativa chromosome 7 clone OJ1341_A08, *** SEQUENCING IN
             PROGRESS ***, in ordered pieces.
ACCESSION    AP003754
VERSION      AP003754.1  GI:14422472
KEYWORDS     HTG; HTGS_PHASE2.
SOURCE       Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1341_A08.
  ORGANISM   Oryza sativa
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1  (sites)
  AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
             clone:OJ1341_A08
  JOURNAL    Published Only in DataBase (2001) In press
REFERENCE    2  (bases 1 to 129052)
  AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE      Direct Submission
  JOURNAL    Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
             Agrobiological Resources, Rice Genome Research Program; Kannondai
             2-1-2, Tsukuba, Ibaraki 305-8602, Japan
             (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
             Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT      The nucleotide sequence of this BAC clone was generated by
             combining Monsanto and RGP-Japan sequencing data.
             NOTE: It currently consists of 1 contigs. Gaps between the contigs
             are represented as runs of N. The order of the pieces is believed
             to be correct as given, however the sizes of the gaps between them
             are based on estimates that have provided by the submitter. This
             sequence will be replaced by the finished sequence as soon as it is
             available and the accession number will be preserved.
             * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .129052
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="7"
                     /clone="OJ1341_A08"
BASE COUNT    35088 a  28842 c  28921 g  36151 t     50 others
ORIGIN


alignment_scores:
            Quality:   69.00              Length:     10
              Ratio:  7.667                Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
```

```
US-09-444-281-36 x AP003754/rev   ..

Align seg 1/1  to reverse of: AP003754  from: 1  to: 129052

      3 ArgTrpProTrpTrpProTrpArgArgLys 12
        |||||||||||||||||||||  |||:::
  53243 CGCTGGCCTTGGTGGCCCTGGACGCGGCGG 53214

seq_name: gb_htg:AC091250

seq_documentation_block:
LOCUS       AC091250   200849 bp    DNA             HTG       11-APR-2001
DEFINITION  Mus musculus chromosome 5 clone RP23-315E2 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION   AC091250
VERSION     AC091250.1  GI:13592171
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200849)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
            Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
            Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200849)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     ---------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            ---------- Project Information
            Center project name: aty
            Center clone name: 315E02
            ---------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197727 bases at least Q40
            Consensus quality: 198423 bases at least Q30
            Consensus quality: 198823 bases at least Q20
            Insert size: 202000; agarose-fp
            Insert size: 200249; sum-of-contigs
            Quality coverage: 9.87x in Q20 bases; agarose-fp
            Quality coverage: 9.95x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2812: contig of 2812 bp in length
            *     2813     2912: gap of unknown length
            *     2913     9301: contig of 6389 bp in length
            *     9302     9401: gap of unknown length
            *     9402    17173: contig of 7772 bp in length
            *    17174    17273: gap of unknown length
            *    17274    26044: contig of 8771 bp in length
            *    26045    26144: gap of unknown length
            *    26145    65929: contig of 39785 bp in length
            *    65930    66029: gap of unknown length
            *    66030   109036: contig of 43007 bp in length
            *   109037   109136: gap of unknown length
            *   109137   200849: contig of 91713 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200849
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-315E2"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .2812
                     /note="assembly_fragment"
     misc_feature    2913. .9301
                     /note="assembly_fragment"
     misc_feature    9402. .17173
                     /note="assembly_fragment"
     misc_feature    17274. .26044
                     /note="assembly_fragment"
     misc_feature    26145. .65929
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    66030. .109036
                     /note="assembly_fragment"
     misc_feature    109137. .200849
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    53444 a  49164 c  49019 g  48614 t    608 others
ORIGIN


alignment_scores:
            Quality:   69.00            Length:     10
              Ratio:  7.667              Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AC091250   ..

 Align seg 1/1  to: AC091250  from: 1  to: 200849

      3 ArgTrpProTrpTrpProTrpArgArgLys 12
        |||||||||||||||||||||||||  :::
 148453 AGGTGGCCTTGGTGGCCCTGGAGGTCTCGG 148482

seq_name: gb_ro:AF210429

seq_documentation_block:
LOCUS       AF210429      456 bp    mRNA            ROD       30-APR-2001
DEFINITION  Mus musculus group X secretory phospholipase A2 (Pla2g10) mRNA,
            complete cds.
ACCESSION   AF210429
VERSION     AF210429.1  GI:12003290
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Morioka,Y., Saiga,A., Yokota,Y., Suzuki,N., Ikeda,M., Ono,T.,
            Nakano,K., Fujii,N., Ishizaki,J., Arita,H. and Hanasaki,K.
  TITLE     Mouse group X secretory phospholipase A2 induces a potent release
            of arachidonic acid from spleen cells and acts as a ligand for the
            phospholipase A2 receptor
  JOURNAL   Arch. Biochem. Biophys. 381 (1), 31-42 (2000)
  MEDLINE   20470496
   PUBMED   11019817
REFERENCE   2  (bases 1 to 456)
  AUTHORS   Morioka,Y., Saiga,A., Yokota,Y., Suzuki,N., Ikeda,M., Ono,T.,
```

```
            Nakano,K., Fujii,N., Ishizaki,J., Arita,H. and Hanasaki,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1999) Shionogi Research Lab, Fukushima-ku Sagisu
            5-12-4, Osaka 553-0002, Japan
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
     gene            1. .456
                     /gene="Pla2g10"
     CDS             1. .456
                     /gene="Pla2g10"
                     /EC_number="3.1.1.4"
                     /note="sPLA2"
                     /codon_start=1
                     /product="group X secretory phospholipase A2"
                     /protein_id="AAG43522.1"
                     /db_xref="GI:12003291"
                     /translation="MLLLLLLLLLGPGPGFSEATRRSHVYKRGLLELAGTLDCVGPRS
                     PMAYMNYGCYCGLGGHGEPRDAIDWCCYHHDCCYSRAQDAGCSPKLDRYPWKCMDHHI
                     LCGPAENKCQELLCRCDEELAYCLAGTEYHLKYLFFPSILCEKDSPKCD"
BASE COUNT       98 a    125 c    129 g    104 t
ORIGIN


alignment_scores:
            Quality:   67.00            Length:     10
              Ratio:  7.444              Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AF210429   ..

 Align seg 1/1  to: AF210429  from: 1  to: 456

      1 IleLeuArgTrpProTrpTrpProTrpArg 10
        :::|||   ||||||||||||||||||||||
    156 CTGTTATTGTGGCCTTGGTGGCCATGGAGA 185

seq_name: gb_ro:AF166097

seq_documentation_block:
LOCUS       AF166097     1040 bp    mRNA            ROD       06-DEC-1999
DEFINITION  Mus musculus group X secreted phospholipase A2 (Pla2g10) mRNA,
            complete cds.
ACCESSION   AF166097
VERSION     AF166097.2  GI:6525307
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Valentin,E., Ghomashchi,F., Gelb,M.H., Lazdunski,M. and Lambeau,G.
  TITLE     On the diversity of secreted phospholipases A(2). Cloning, tissue
            distribution, and functional expression of two novel mouse group II
            enzymes
  JOURNAL   J. Biol. Chem. 274 (44), 31195-31202 (1999)
  MEDLINE   20002639
  REMARK    Erratum:[[published erratum appears in J Biol Chem 2000 Jan
            21;275(3):2246]]
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Valentin,E., Ghomashchi,F., Gelb,M.H., Lazdunski,M. and Lambeau,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1999) IPMC, CNRS, 660, route des Lucioles,
            Valbonne 06560, France
REFERENCE   3  (bases 1 to 1040)
  AUTHORS   Valentin,E., Ghomashchi,F., Gelb,M.H., Lazdunski,M. and Lambeau,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1999) IPMC, CNRS, 660, route des Lucioles,
            Valbonne 06560, France
  REMARK    Sequence update by submitter
COMMENT     On Dec 6, 1999 this sequence version replaced gi:6164695.
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="16"
     gene            1. .1040
                     /gene="Pla2g10"
     CDS             175. .630
                     /gene="Pla2g10"
                     /EC_number="3.1.1.4"
                     /note="Ca2+-dependent secreted phospholipase A2"
                     /codon_start=1
                     /product="group X secreted phospholipase A2"
                     /protein_id="AAF04498.2"
                     /db_xref="GI:6525308"
                     /translation="MLLLLLLLLLGPGPGFSEATRRSHVYKRGLLELAGTLDCVGPRS
                     PMAYMNYGCYCGLGGHGEPRDAIDWCCYHHDCCYSRAQDAGCSPKLDRYPWKCMDHHI
                     LCGPAENKCQELLCRCDEELAYCLAGTEYHLKYLFFPSILCEKDSPKCN"
BASE COUNT      244 a    258 c    277 g    261 t
ORIGIN


alignment_scores:
            Quality:   67.00            Length:     10
              Ratio:  7.444              Gaps:      0
 Percent Similarity:  90.000   Percent Identity: 80.000

alignment_block:
 US-09-444-281-36 x AF166097   ..

 Align seg 1/1  to: AF166097  from: 1  to: 1040

      1 IleLeuArgTrpProTrpTrpProTrpArg 10
        :::|||   ||||||||||||||||||||||
    330 CTGTTATTGTGGCCTTGGTGGCCATGGAGA 359

seq_name: gb_pl:ATPHYB

seq_documentation_block:
LOCUS       ATPHYB       3850 bp    mRNA            PLN       13-SEP-1994
DEFINITION  Arabidopsis thaliana phyB mRNA for phytochrome.
ACCESSION   X17342
VERSION     X17342.1  GI:16422
KEYWORDS    photoreceptor; phytochrome.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3850)
  AUTHORS   Sharrock,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1989) Sharrock, R., Montana State University,
            Department of Biology, Bozeman, MT 59717 USA
REFERENCE   2  (bases 1 to 3850)
  AUTHORS   Sharrock,R.A. and Quail,P.H.
  TITLE     Novel phytochrome sequences in Arabidopsis thaliana: structure,
            evolution, and differential expression of a plant regulatory
            photoreceptor family
  JOURNAL   Genes Dev. 3 (11), 1745-1757 (1989)
  MEDLINE   90108670
FEATURES             Location/Qualifiers
     source          1. .3850
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /dev_stage="3 week old green tissue"
                     /tissue_type="leaf."
                     /clone_lib="lambda gt10"
                     /clone="A7-5"
     CDS             93. .3611
                     /note="phyB photoreceptor"
```

```
                     /codon_start=1
                     /protein_id="CAA35222.1"
                     /db_xref="GI:16423"
                     /db_xref="SWISS-PROT:P14713"
                     /translation="MVSGVGGSGGGRGGGRGGEEEPSSSHTPNNRRGGEQAQSSGTKS
                     LRPRSNTESMSKAIQQYTVDARLHAVFEQSGESGKSFDYSQSLKTTTYGSSVPEQQIT
                     AYLSRIQRGGYIQPFGCMIAVDESSFRIIGYSENAREMLGIMPQSVPTLEKPEILAMG
                     TDVRSLFTSSSSILLERAFVAREITLLNPVWIHSKNTGKPFYAILHRIDVGVVIDLEP
                     ARTEDPALSIAGAVQSQKLAVRAISQLQALPGGDIKLLCDTVVESVRDLTGYDRVMVY
                     KFHEDEHGEVVAESKRDDLEPYIGLHYPATDIPQASRFLFKQNRVRMIVDCNATPVLV
                     VQDDRLTQSMCLVGSTLRAPHGCHSQYMANMGSIASLAMAVIINGNEDDGSNVASGRS
                     SMRLWGLVVCHHTSSRCIPFPLRYACEFLMQAFGLQLNMELQLALQMSEKRVLRTQTL
                     LCDMLLRDSPAGIVTQSPSIMDLVKCDGAAFLYHGKYYPLGVAPSEVQIKDVVEWLLA
                     NHADSTGLSTDSLGDAGYPGAAALGDAVCGMAVAYITKRDFLFWFRSHTAKEIKWGGA
                     KHHPEDKDDGQRMHPRSSFQAFLEVVKSRSQPWETAEMDAIHSLQLILRDSFKESEAA
                     MNSKVVDGVVQPCRDMAGEQGIDELGAVAREMVRLIETATVPIFAVDAGGCINGWNAK
                     IAELTGLSVEEAMGKSLVSDLIYKENEATVNKLLSRALRGDEEKNVEVKLKTFSPELQ
                     GKAVFVVVNACSSKDYLNNIVGVCFVGQDVTSQKIVMDKFINIQGDYKAIVHSPNPLI
                     PPIFAADENTCCLEWNMAMEKLTGWSRSEVIGKMIVGEVFGSCCMLKGPDALTKFMIV
                     LHNAIGGQDTDKFPFPFFDRNGKFVQALLTANKRVSLEGKVIGAFCFLQIPSPELQQA
                     LAVQRRQDTECFTKAKELAYICQVIKNPLSGMRFANSLLEATDLNEDQKQLLETSVSC
                     EKQISRIVGDMDLESIEDGSFVLKREEFFLGSVINAIVSQAMFLLRDRGLQLIRDIPE
                     EIKSIEVFGDQIRIQQLLAEFLLSIIRYAPSQEWVEIHLSQLSKQMADGFAAIRTEFR
                     MACPGEGLPPELVRDMFHSSRWTSPEGLGLSVCRKILKLMNGEVQYIRESERSYFLII
                     LELPVPRKRPLSTASGSGDMMLMMPY"
     polyA_site      3850
                     /note="polyA site"
BASE COUNT     1021 a    734 c   1013 g   1082 t
ORIGIN


alignment_scores:
            Quality:   64.50             Length:     11
              Ratio:  6.450               Gaps:      1
 Percent Similarity:  90.909   Percent Identity: 81.818

alignment_block:
 US-09-444-281-36 x ATPHYB   ..

 Align seg 1/1  to: ATPHYB  from: 1  to: 3850

       3 ArgTrpProTrp...TrpProTrpArgArgLys 12
         |||||||||||   ||||||||||||||:::
     119 CGGTGGCCGTGGCGGTGGCCGTGGCGGAGAAGA 151

seq_name: gb_pl:ATHPHTOCHB

seq_documentation_block:
LOCUS        ATHPHTOCHB   6509 bp    DNA             PLN       14-JUN-1993
DEFINITION   Arabidopsis thaliana phytochrome b gene, exons 2 and 3.
ACCESSION    L09262
VERSION      L09262.1  GI:166793
KEYWORDS     photoreceptor; phytochrome B.
SOURCE       Arabidopsis thaliana (strain Landsberg erecta) DNA.
  ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1  (bases 1 to 6509)
  AUTHORS    Reed,J.W., Nagpal,P., Poole,D.S., Furuya,M. and Chory,J.
  TITLE      Mutations in the gene for the red/far red light receptor
             phytochrome B alter cell elongation and physiological responses
             throughout Arabidopsis development
  JOURNAL    Plant Cell 5, 147-157 (1993)
  MEDLINE    93200802
FEATURES              Location/Qualifiers
     source           1. .6509
                      /organism="Arabidopsis thaliana"
                      /strain="Landsberg erecta"
                      /db_xref="taxon:3702"
     5'UTR            2201. .2298
     intron           4463. .4550
                      /number=1
     exon             4551. .5358
                      /number=2
     intron           5359. .5448
     exon             5449. .5739
                      /number=3
     intron           5740. .6115
                      /number=3
     3'UTR            6369. .6509
BASE COUNT     1823 a   1192 c   1459 g   2035 t
ORIGIN


alignment_scores:
            Quality:   64.50             Length:     11
              Ratio:  6.450               Gaps:      1
 Percent Similarity:  90.909   Percent Identity: 81.818

alignment_block:
 US-09-444-281-36 x ATHPHTOCHB   ..

 Align seg 1/1  to: ATHPHTOCHB  from: 1  to: 6509

       3 ArgTrpProTrp...TrpProTrpArgArgLys 12
         |||||||||||   ||||||||||||||:::
    2325 CGGTGGCCGTGGCGGTGGCCGTGGCGGAGAAGA 2357

seq_name: gb_pl:AC005724

seq_documentation_block:
LOCUS        AC005724    86671 bp    DNA             PLN       05-APR-2000
DEFINITION   Arabidopsis thaliana chromosome II section 109 of 255 of the
             complete sequence. Sequence from clones MSF3.
ACCESSION    AC005724 AE002093
VERSION      AC005724.2  GI:6598484
KEYWORDS     HTG.
SOURCE       thale cress.
  ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1  (bases 1 to 86671)
  AUTHORS    Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
             Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
             Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
             Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
             Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
             Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
             Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
             Venter,J.C.
  TITLE      Sequence and analysis of chromosome 2 of the plant Arabidopsis
             thaliana
  JOURNAL    Nature 402 (6763), 761-768 (1999)
  MEDLINE    20083487
   PUBMED    10617197
REFERENCE    2  (bases 1 to 86671)
  AUTHORS    Lin,X.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
             Medical Center Dr., Rockville, MD 20850, USA
COMMENT      On Dec 17, 1999 this sequence version replaced gi:4185128.
             The sequence and annotation of chromosome 2 were merged from those
             of the individual clones on this chromosome after removing
             overlaps. For detailed information, please see the TIGR web site
             (http://www.tigr.org/tdb/at/at.html).

             Genes were identified by a combination of three methods: Gene
             prediction programs including GRAIL
             (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
             University of Washington), Genscan (Chris Burge,
             http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
             (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
             complete sequence against a peptide database and plant EST
             databases at TIGR, and manual curations based on those analyses.
```

Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source 1. .86671
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

misc_feature 1. .>86671
/note="Sequence from clone MSF3"

repeat_region complement(345. .409)
/rpt_family="(GAAAA)n"

mRNA join(<1030. .2192,2354. .2494,2618. .2896)
/gene="At2g18630"

gene <1030. .2896
/gene="At2g18630"
/note="MSF3.1; unusual splice site at second intron; GA instead of conserved GT at donor site"

CDS join(1030. .2192,2354. .2369)
/gene="At2g18630"
/note="unknown protein"
/codon_start=1
/protein_id="AAD08932.1"
/db_xref="GI:4185129"
/translation="MGGKSSKSKKNVEFGSPSTPVQIKINSEYTEHLSSYERACSEDP
KLESFDSALHERTNRVINKLASGVEIKSLSFDSLREVTQCLLDMNQDVVKVILQDKED
IWNNQDLFSLVNLYFESTAKTMDFCSELENCLNRARRSQVIIQFAVNQFEEENEDKEN
RKYEKTLEELKRFKVAGEPFTKEFFALFDLVYKQQVMMLEELHKLKRKLDKRLRNIKT
WRRVSNMVFVTAFVSVLIFSVVAAAVAAPPVVAAIAGALAVPVGSVGKWCNTLWTKYE
KVVRGQKEIITSIRIGTYISVKEMDNISILVRKVEVEIESLLKKAEFAITEEKEVRLA
IDEIKKKLDVFTETIEELGEHAGKYCSDVTKARTVILQRIIRYPAGSPKDEAPWTEMM
"

mRNA complement(3613. .4761)
/gene="At2g18640"

gene complement(3613. .4761)
/gene="At2g18640"
/note="MSF3.2; contains GB:L22347"

CDS complement(3627. .4745)
/gene="At2g18640"
/codon_start=1
/product="putative geranylgeranyl pyrophosphate synthase"
/protein_id="AAD08933.1"
/db_xref="GI:4185130"
/translation="MEAQNIFLYLLIVFLSLHFVFTTLKGRLSPANTRRLIRLLHIPI
KSPVAAAIFARKDTREFLDSSIKLVNEEDDFGFSFDFKPYMISKAETINRALDEAIPL
IEPLNIHKAMRYAILAGGKRVRPILCLAACELVGGEERLAIQAACAVEMIHTMSLIKD
DLPCMDNDDLRRGKPTTHKVFGESVAILSGGALLALAFEHLTEADVSSKKMVRAVKEL
AKSIGTKGLVAGQAKDLSSEGLEQNDVGLEDLEYIHVHKTGSLLEASAVIGAVIGGGT
EKEIEKVRNFARCIGLLFQVVDDILDETKSSEELGKTAGKDKVAGKLTYPKVIGVEKS
KEFVEKLKRDAREHLQGFDSDKVKPLIALTNFIANRNH"

mRNA complement(<6733. .>8004)
/gene="At2g18650"

gene complement(<6733. .>8004)
/gene="At2g18650"
/note="MSF3.3"

CDS complement(6733. .8004)
/gene="At2g18650"
/codon_start=1
/product="putative C3HC4-type RING zinc finger protein"
/protein_id="AAD08934.1"
/db_xref="GI:4185131"
/translation="MNKILPEMKSTQNLISSSPPPPLIPLKSNTSLSNLNSKITPNIL
LIIIILSIIFFISGLLHILVKFLLTPSRESREDYFDNVTALQGQLQQLFNLHDSGVDQ
SLIDTLPVFHYKSIVGLKISPFDCPVCLCEFETEDKLRLLPKCSHAFHVECIDTWLLS
HSTCPLCRSNLLSGFSSHHNLSSSYLLVLESEQSSRDMVPVLESNSQLGYDVNNDSES
TRIRSGRKSCDPDGDMDGLDEKVVPLEVKLGKFRNIDHVGEGSDQKKNSISGNSKNVD
GRRCLSMGSYEYIMDQEATLKVHVSTKKLSGKDRVPSHRTVMSECGFDPTVKGIEKSV
VERESFSLSKIWLRGKKEKQKGTSARDSDCSFVSSSSLRFPNHRIPPEESLKSENSES
LETKTPSFARRTMHWLAGRQNKIVQPSTSNV"

repeat_region 9406. .9449
/rpt_family="(GA)n"

repeat_region 10570. .10636
/rpt_family="(TA)n"

mRNA complement(join(<11029. .11303,11403. .>11508))
/gene="At2g18660"

gene complement(<11029. .>11508)
/gene="At2g18660"
/note="MSF3.4; predicted by genscan"

CDS complement(join(11029. .11303,11403. .11508))
/gene="At2g18660"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD08935.1"
/db_xref="GI:4185132"
/translation="MAVKFVVVMIVFAQILAPIAEAAQGKAVYYDPPYTRSACYGTQR
ETLVVGVKNNLWQNGRACGRRYRVRCIGATYNFDRACTGRTVDVKVVDFCREPCNGDL
NLSRDAFRVIANTDAGNIRVVYTP"

repeat_region 11334. .11401
/rpt_family="(TA)n"

mRNA <13342. .>13887
/gene="At2g18670"

gene <13342. .>13887
/gene="At2g18670"
/note="MSF3.5"

CDS 13342. .13887
/gene="At2g18670"
/codon_start=1
/product="putative C3HC4-type RING zinc finger protein"
/protein_id="AAD08936.1"
/db_xref="GI:4185133"
/translation="MPPTNNYRISGEPPSTTPSHPPPKPKTRILSLFLVGVIMFSIFF
LFLVLIGIASVLILPLLLSSLHRHHRRRRRNRRQESSDGLSSRFVKKLPQFKFSEPST
YTRYESDCVVCFDGFRQGQWCRNLPGCGHVFHRKCVDTWLLKASTCPICRARVRLWEE
DPQEGELWRCFGHRRSSLLDL"

mRNA 14514. .15636
/gene="At2g18680"

gene 14514. .15636
/gene="At2g18680"
/note="MSF3.6"

CDS 14653. .15516
/gene="At2g18680"
/note="unknown protein"
/codon_start=1
/protein_id="AAD08937.1"
/db_xref="GI:4185134"
/translation="MFSVLVFPLLLNCLVYLFNAIAIKPEITNLILESSLLPMTDPNT
PEFAAHLMRVFVDFRQFVNSLYIFIAVSSIINLLSTLVMVHASALTHKDDSFEIKDFP
ILTLKYWKGPLVTNFYIVLFSLGYWFLFFIVLFSIVFFSTKLDSLAAKSRALFIVFAV
FESYLAIVWNLSMVISILEDTYGIQALGKAAKIVKGMKPKLFLLNLFFGLLSFGLVQI
LRLVDWSSSFSVTLTTGLVLMSSVFVVRMFQLVTYTVAYFQCKSLQGQDSLRDVEYTK
LSSTTLMGGLP"

mRNA 17474. .18643

```
                     /gene="At2g18690"
     gene            17474. .18643
                     /gene="At2g18690"
                     /note="MSF3.7"
     CDS             17523. .18491
                     /gene="At2g18690"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAD08938.1"
                     /db_xref="GI:4185135"

alignment_scores:
            Quality:   64.50              Length:     11
              Ratio:  6.450                Gaps:      1
 Percent Similarity:  90.909   Percent Identity: 81.818

alignment_block:
 US-09-444-281-36 x AC005724   ..

 Align seq 1/1  to: AC005724  from: 1  to: 86671

       3 ArgTrpProTrp...TrpProTrpArgArgLys 12
         ||||||||||||   ||||||||||||||:::
   59978 CGGTGGCCGTGGCGGTGGCCGTGGCGGAGAAGA 60010

seq_name: gb_pl:AP003141

seq_documentation_block:
LOCUS       AP003141   141275 bp    DNA              PLN       01-FEB-2001
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0002B05,
            complete sequence.
ACCESSION   AP003141
VERSION     AP003141.1  GI:12641875
KEYWORDS    HTG.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0002B05.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 141275)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0002B05
  JOURNAL   Published Only in DataBase (2001) In press
REFERENCE   2  (bases 1 to 141275)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from T7 to SP6 of the PAC clone.
FEATURES             Location/Qualifiers
     source          1. .141275
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
                     /clone="P0002B05"
BASE COUNT    37539 a  33263 c  31475 g  38998 t
ORIGIN


alignment_scores:
            Quality:   64.50              Length:     11
              Ratio:  6.450                Gaps:      1
 Percent Similarity:  90.909   Percent Identity: 81.818

alignment_block:
 US-09-444-281-36 x AP003141.rev   ..

 Align seq 1/1  to reverse of: AP003141  from: 1  to: 141275

       3 ArgTrpProTrpTrpPro...TrpArgArgLys 12
         ||||||||||||||||||   ||||||||||:::
   68277 CGATGGCCTTGGTGGCCTCGCTGGCGGCGACGA 68245

seq_name: gb_ro:MMA2IXCOA

seq_documentation_block:
LOCUS       MMA2IXCOA   19479 bp    DNA              ROD       08-APR-1994
DEFINITION  M.musculus alpha2 (IX) collagen gene, complete CDS.
ACCESSION   Z22923
VERSION     Z22923.1  GI:311949
KEYWORDS    alpha2 (IX) collagen.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 19479)
  AUTHORS   Perala,M., Elima,K., Metsaranta,M., Rosati,R., de Crombrugghe,B.
            and Vuorio,E.
  TITLE     The exon structure of the mouse alpha 2(IX) collagen gene shows
            unexpected divergence from the chick gene
  JOURNAL   J. Biol. Chem. 269 (7), 5064-5071 (1994)
  MEDLINE   94148964
REFERENCE   2  (bases 1 to 19479)
  AUTHORS   Perala,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-1993) Merja Perala, Medical Biochemistry,
            University of Turku, Kiinamyllynkatu 10, 20520 Turku, Finland
FEATURES             Location/Qualifiers
     source          1. .19479
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="cMP9A2"
                     /clone_lib="Mouse genomic cosmid library (Clonetech)"
     CAAT_signal     2600. .2604
     exon            2677. .3131
                     /gene="alpha2 (IX) collagen"
                     /note="alternative transcription start site= 2558"
                     /label=ex1
                     /number=1
                     /evidence=experimental
     mRNA            join(2677. .3131,4311. .4385,5646. .5681,5768. .5830,
                     7048. .7101,7214. .7249,7347. .7370,7630. .7683,8010. .8063,
                     8168. .8215,8344. .8400,8485. .8538,8828. .8881,9268. .9321,
                     9413. .9466,9572. .9625,12979. .13032,13393. .13446,
                     13672. .13725,14530. .14574,14967. .15020,15284. .15337,
                     15543. .15596,15674. .15745,15897. .15932,16069. .16113,
                     16207. .16239,16375. .16521,16756. .16810,17086. .17274,
                     17569. .17646,17912. .18617)
                     /gene="alpha2 (IX) collagen"
     gene            2677. .18617
                     /gene="alpha2 (IX) collagen"
     CDS             join(3060. .3131,4311. .4385,5646. .5681,5768. .5830,
                     7048. .7101,7214. .7249,7347. .7370,7630. .7683,8010. .8063,
                     8168. .8215,8344. .8400,8485. .8538,8828. .8881,9268. .9321,
                     9413. .9466,9572. .9625,12979. .13032,13393. .13446,
                     13672. .13725,14530. .14574,14967. .15020,15284. .15337,
                     15543. .15596,15674. .15745,15897. .15932,16069. .16113,
                     16207. .16239,16375. .16521,16756. .16810,17086. .17274,
                     17569. .17646,17912. .18111)
                     /gene="alpha2 (IX) collagen"
                     /codon_start=1
                     /product="alpha2 (IX) collagen"
                     /protein_id="CAA80503.1"
                     /db_xref="GI:311950"
                     /db_xref="MGD:MGI:88466"
                     /db_xref="SPTREMBL:Q07643"
                     /translation="MTAVPAPRSLFVLLQVLWLALAQIRGPPGEPGPPGPPGPPGVPG
                     SDGIDGDKGPPGKVGPPGSKGEPGKPGPDGPDGKPGIDGLMGAKGEPGPVGTPGVKGQ
                     PGLPGPPGLPGPGFAGPPGPPGPVGLPGEIGTPGPKGDPGPEGPSGPPGPPGKPGRPG
                     TIQGLEGSADFLCPTNCPAGVKGPQGLQGVKGHPGKRGILGDPGRQGKPGPKGDVGAS
```

GEQGIPGPPGPQGIRGYPGMAGPKGEMGPRGYKGMVGSIGAAGPPGEEGPRGPPGEAG
EKGDVGSQGARGPQGITGPKGITGPPGIDGKDGTPGIPGMKGSAGQVGRPGSPGHQGL
AGVPGQPGTKGGPGDKGEPGQQGLPGVSGPPGKEGEPGPRGEIGPQGIMGQKGDQGER
GPVGQPGPQGRQGPKGEQGPPGIPGPQGLPGIKGDKGSPGKTGPRGGVGDPGVAGLPG
EKGEKGQSGEPGLKGQQGVRGETGYPGPSGDIGAPGVQGYPGLPGPRGLVGDRGVPGQ
PGRQGVVGRAASDQHIVDVVLKMIQEQLAEVAVSAKREALGAAGMVGLPGPPGPPGYP
GKQGPNGHPGPRGIPGIVGAVGQIGNTGPKGKRGEKGDRGEMGHGHPGMPGPPGIPGL
PGRPGQAINGKDGDRGSPGAPGEAGRPGRPGPVGLPGFCEPAACLGASAYTSARLTEP
GSIKGP"

exon	4311. .4385	/gene="alpha2 (IX) collagen"
		/label=ex2
		/number=2
		/evidence=experimental
exon	5646. .5681	/gene="alpha2 (IX) collagen"
		/label=ex3
		/number=3
		/evidence=experimental
exon	5768. .5830	/gene="alpha2 (IX) collagen"
		/label=ex4
		/number=4
		/evidence=experimental
exon	7048. .7101	/gene="alpha2 (IX) collagen"
		/label=ex5
		/number=5
		/evidence=experimental
exon	7214. .7249	/gene="alpha2 (IX) collagen"
		/label=ex6
		/number=6
		/evidence=experimental
exon	7347. .7370	/gene="alpha2 (IX) collagen"
		/label=ex7
		/number=7
		/evidence=experimental
exon	7630. .7683	/gene="alpha2 (IX) collagen"
		/label=ex8
		/number=8
		/evidence=experimental
exon	8010. .8063	/gene="alpha2 (IX) collagen"
		/label=ex9
		/number=9
		/evidence=experimental
exon	8168. .8215	/gene="alpha2 (IX) collagen"
		/label=ex10
		/number=10
		/evidence=experimental
exon	8344. .8400	/gene="alpha2 (IX) collagen"
		/label=ex11
		/number=11
		/evidence=experimental
exon	8485. .8538	/gene="alpha2 (IX) collagen"
		/label=ex12
		/number=12
		/evidence=experimental
exon	8828. .8881	/gene="alpha2 (IX) collagen"
		/label=ex13
		/number=13
		/evidence=experimental
exon	9268. .9321	/gene="alpha2 (IX) collagen"
		/label=ex14
		/number=14
		/evidence=experimental
exon	9413. .9466	/gene="alpha2 (IX) collagen"
		/label=ex15
		/number=15
		/evidence=experimental
exon	9572. .9625	/gene="alpha2 (IX) collagen"
		/label=ex16
		/number=16
		/evidence=experimental
exon	12979. .13032	/gene="alpha2 (IX) collagen"
		/label=ex17
		/number=17
		/evidence=experimental
exon	13393. .13446	/gene="alpha2 (IX) collagen"
		/label=ex18
		/number=18
		/evidence=experimental
exon	13672. .13725	/gene="alpha2 (IX) collagen"
		/label=ex19
		/number=19
		/evidence=experimental
exon	14530. .14574	/gene="alpha2 (IX) collagen"
		/label=ex20
		/number=20
		/evidence=experimental
exon	14967. .15020	/gene="alpha2 (IX) collagen"
		/label=ex21
		/number=21
		/evidence=experimental
exon	15284. .15337	/gene="alpha2 (IX) collagen"
		/label=ex22
		/number=22
		/evidence=experimental
exon	15543. .15596	/gene="alpha2 (IX) collagen"
		/label=ex23
		/number=23
		/evidence=experimental
exon	15674. .15745	/gene="alpha2 (IX) collagen"
		/label=ex24
		/number=24
		/evidence=experimental
exon	15897. .15932	/gene="alpha2 (IX) collagen"
		/label=ex25
		/number=25
		/evidence=experimental
exon	16069. .16113	/gene="alpha2 (IX) collagen"
		/label=ex26
		/number=26
		/evidence=experimental
exon	16207. .16239	

alignment_scores:
Quality: 64.00 Length: 9
Ratio: 8.000 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 77.778

alignment_block:
US-09-444-281-36 x MMA2IXCOA/rev ..

Align seg 1/1 to reverse of: MMA2IXCOA from: 1 to: 19479

```
      4 TrpProTrpTrpProTrpArgArgLys 12
        ||||||||||||||||||||    :::
   4355 TGGCCCTGGTGGCCCTGGAGGACCCGG 4329

seq_name: gb_htg:AL356097_0

seq_documentation_block:
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AL356097 Accession AL356097
     Fragment Name      Begin     End
     AL356097_0             1    110000
     AL356097_1        100001    210000
     AL356097_2        200001    310000
     AL356097_3        300001    355026
LOCUS       AL356097   355026 bp    DNA             HTG       07-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-180A14, *** SEQUENCING IN
            PROGRESS ***, 38 unordered pieces.
ACCESSION   AL356097 AC058797
VERSION     AL356097.11  GI:11414557
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355026)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 15, 2001 this sequence version replaced gi:8077023
            gi:9797422.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA180A14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 43% of reads
            Sequencing vector: plasmid; L08752; 56% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 333302 bases at least Q40
            Consensus quality: 341682 bases at least Q30
            Consensus quality: 346727 bases at least Q20
            Insert size: 351326; sum-of-contigs
            Insert size: 176896; 2.8% error; agarose-fp
            Quality coverage: 3.95x in Q20 bases; sum-of-contigs Quality
            coverage: 8.06x in Q20 bases; agarose-fp
            --------------
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    16446: contig of 16446 bp in length
            *    16447 16546: gap of      100 bp
            *    16547    22746: contig of 6200 bp in length
            *    22747 22846: gap of      100 bp
            *    22847    45482: contig of 22636 bp in length
            *    45483 45582: gap of      100 bp
            *    45583    49251: contig of 3669 bp in length
            *    49252 49351: gap of      100 bp
            *    49352    53239: contig of 3888 bp in length
            *    53240 53339: gap of      100 bp
            *    53340    58178: contig of 4839 bp in length
            *    58179 58278: gap of      100 bp
            *    58279    71754: contig of 13476 bp in length
            *    71755 71854: gap of      100 bp
            *    71855    76641: contig of 4787 bp in length
            *    76642 76741: gap of      100 bp
            *    76742    80415: contig of 3674 bp in length
            *    80416 80515: gap of      100 bp
            *    80516    89902: contig of 9387 bp in length
            *    89903 90002: gap of      100 bp
            *    90003    93661: contig of 3659 bp in length
            *    93662 93761: gap of      100 bp
            *    93762   101785: contig of 8024 bp in length
            *   101786 101885: gap of      100 bp
            *   101886   109333: contig of 7448 bp in length
            *   109334 109433: gap of      100 bp
            *   109434   121288: contig of 11855 bp in length
            *   121289 121388: gap of      100 bp
            *   121389   128767: contig of 7379 bp in length
            *   128768 128867: gap of      100 bp
            *   128868   131917: contig of 3050 bp in length
            *   131918 132017: gap of      100 bp
            *   132018   138795: contig of 6778 bp in length
            *   138796 138895: gap of      100 bp
            *   138896   144660: contig of 5765 bp in length
            *   144661 144760: gap of      100 bp
            *   144761   156587: contig of 11827 bp in length
            *   156588 156687: gap of      100 bp
            *   156688   158939: contig of 2252 bp in length
            *   158940 159039: gap of      100 bp
            *   159040   161353: contig of 2314 bp in length
            *   161354 161453: gap of      100 bp
            *   161454   165170: contig of 3717 bp in length
            *   165171 165270: gap of      100 bp
            *   165271   168479: contig of 3209 bp in length
            *   168480 168579: gap of      100 bp
            *   168580   172188: contig of 3609 bp in length
            *   172189 172288: gap of      100 bp
            *   172289   177084: contig of 4796 bp in length
            *   177085 177184: gap of      100 bp
            *   177185   199876: contig of 22692 bp in length
            *   199877 199976: gap of      100 bp
            *   199977   222186: contig of 22210 bp in length
            *   222187 222286: gap of      100 bp
            *   222287   240235: contig of 17949 bp in length
            *   240236 240335: gap of      100 bp
            *   240336   263715: contig of 23380 bp in length
            *   263716 263815: gap of      100 bp
            *   263816   269270: contig of 5455 bp in length
            *   269271 269370: gap of      100 bp
            *   269371   271745: contig of 2375 bp in length
            *   271746 271845: gap of      100 bp
            *   271846   277024: contig of 5179 bp in length
            *   277025 277124: gap of      100 bp
            *   277125   280037: contig of 2913 bp in length
            *   280038 280137: gap of      100 bp
            *   280138   307328: contig of 27191 bp in length
            *   307329 307428: gap of      100 bp
            *   307429   313985: contig of 6557 bp in length
            *   313986 314085: gap of      100 bp
            *   314086   330085: contig of 16000 bp in length
            *   330086 330185: gap of      100 bp
            *   330186   334887: contig of 4702 bp in length
            *   334888 334987: gap of      100 bp
            *   334988   355026: contig of 20039 bp in length.
FEATURES             Location/Qualifiers
     source          1. .355026
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-180A14"
                     /clone_lib="RPCI-11.1"
```

```
     misc_feature    1. .16446
                     /note="assembly_fragment:02644
                     fragment_chain:1"
     misc_feature    16547. .22746
                     /note="assembly_fragment:02589
                     fragment_chain:1"
     misc_feature    22847. .45482
                     /note="assembly_fragment:03035
                     fragment_chain:1"
     misc_feature    45583. .49251
                     /note="assembly_fragment:00032"
     misc_feature    49352. .53239
                     /note="assembly_fragment:00060"
     misc_feature    53340. .58178
                     /note="assembly_fragment:00105"
     misc_feature    58279. .71754
                     /note="assembly_fragment:00250"
     misc_feature    71855. .76641
                     /note="assembly_fragment:00381"
     misc_feature    76742. .80415
                     /note="assembly_fragment:00389"
     misc_feature    80516. .89902
                     /note="assembly_fragment:00591"
     misc_feature    90003. .93661
                     /note="assembly_fragment:00750"
     misc_feature    93762. .101785
                     /note="assembly_fragment:00754"
     misc_feature    101886. .109333
                     /note="assembly_fragment:00819"
     misc_feature    109434. .121288
                     /note="assembly_fragment:00889"
     misc_feature    121389. .128767
                     /note="assembly_fragment:00932"
     misc_feature    128868. .131917
                     /note="assembly_fragment:00972"
     misc_feature    132018. .138795
                     /note="assembly_fragment:01033"
     misc_feature    138896. .144660
                     /note="assembly_fragment:01363"
     misc_feature    144761. .156587
                     /note="assembly_fragment:01542"
     misc_feature    156688. .158939
                     /note="assembly_fragment:01630"
     misc_feature    159040. .161353
                     /note="assembly_fragment:01684"
     misc_feature    161454. .165170
                     /note="assembly_fragment:01912"
     misc_feature    165271. .168479
                     /note="assembly_fragment:01914"
     misc_feature    168580. .172188
                     /note="assembly_fragment:02231"
     misc_feature    172289. .177084
                     /note="assembly_fragment:02455"
     misc_feature    177185. .199876
                     /note="assembly_fragment:02553"
     misc_feature    199977. .222186
                     /note="assembly_fragment:02605"
     misc_feature    222287. .240235
                     /note="assembly_fragment:02624"
     misc_feature    240336. .263715
                     /note="assembly_fragment:03030"

alignment_scores:
           Quality:   64.00          Length:      9
             Ratio:  7.111            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
 US-09-444-281-36 x AL356097_0/rev   ..

 Align seg 1/1  to reverse of: AL356097_0  from: 1  to: 110000

       4 TrpProTrpTrpProTrpArgArgLys 12
                 ||||||||||||||||||:::::
    8764 TGGCCCTGGTGGCCTTGGAGGGAGGAG 8738

seq_name: gb_pl:AP001366

seq_documentation_block:
LOCUS       AP001366   146081 bp    DNA             PLN       12-AUG-2000
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0469E09.
ACCESSION   AP001366 BA000010
VERSION     AP001366.1  GI:7228436
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0469E09.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 146081)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0469E09
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 146081)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October 1998 version). The genomic sequence was
            searched against the non-redundant database NRP (PIR, SWISSPROT,
            GENPEPT, PDB) from MAFF DNA bank and the cDNA sequence database at
            RGP. Protein similarities of the cording regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN 2.0 with the corresponding DDBJ accession
            no. and RGP clone ID.
            Detailed information on overlap and assemble quality together with
            annotation of this entry at
            http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.
FEATURES              Location/Qualifiers
     source           1. .146081
                      /organism="Oryza sativa"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /chromosome="1"
                      /clone="P0469E09"
     CDS              join(5902. .5980,6229. .6272,8134. .8295)
                      /note="hypothetical protein"
                      /codon_start=1
                      /protein_id="BAA92397.1"
                      /db_xref="GI:7228437"
                      /translation="MGAVKRSSFLLVVVVFALLLLTSMAAGGRKMLINKHQVQSMETS
                      DDESMHQGQEDDEMLAMVHERILRQVKTNDYGTYDPTPTMAKPHAKEIPN"
     LTR              9673. .9989
                      /note="5' LTR"
     CDS              join(10528. .12588,12673. .13083,13175. .13473,13550. .13808,
                      13901. .14083,14241. .14306,14743. .15045,15216. .15479)
                      /note="EST C28952(C62945) corresponds to a region of the
                      predicted gene.
                      Similar to maize transposon MuDR mudrA protein isolog
                      (AC003981)"
                      /codon_start=1
                      /protein_id="BAA92398.1"
                      /db_xref="GI:7228438"
                      /translation="MTSKGIDRPAERTFQSIYSWLLRGFRIDQEVYTMSVSVVVSRAT
                      EGYFWELMPMDSTAAWRRYVEMAFERSWPLVIFVSVQEKDINVSMQTEDVEGPINAGD
                      VVGPSMQNEENQPREEQAMGMADEGERVGIIVDEMEREDSDNEEADDDASSDEEGDVM
                      ATDWANEDFSGLVILEGDHVPWEYKENEVIDGARYAHKDEMKEAVKHWAVSLQREFRV
                      VKSTNYVYEVRCMKEDCPWRVHAYKGKWNDYWKVSIVTEHKCYLQGVEKYHRNITSAF
                      VASEMYSSVVGNIGFEPKSIIRHIENKFKYTISYVKACRAKQKIIEMRYGTFEASYDN
```

LPRLLATIAQRNNNTYYDLHTFTSVDDRTKSVLQRAFFSLGACINAFVHCRPVLCIDG
TFMTGKYRGQILTAIGCDGNNQVVPMAFAFVESENTESWYWFLERVHIAVVRMRPNVC
LIHDRHAGMLRAIDYLQNGWDEKRLPAKWPDVRSRWCMRHMGANFYKQFKNKHLMELF
KRLCAKNQEKKFNELWDKLDELTTKQTDEQSRRPQVEGDEPPIPLGALHDDPPTMRRR
SGSSIRNFTQWIENEPTEKWSLLFDTDGSRYGIMTTNLAEVYNWVMRGVRVLPLVAIV
EFILHGTQAYFRDRYKKIGPSMADNNIVFGFEQVMRLPHLGPANTLPPYSTVGPSKAW
LLQFTADLLHPDADNNSVRRSLEAYLLWLFGWVMFTSTHGHAVDFRLVHYARSIVDAQ
PQDVPQWSWGSAVLAATYRALCEACTKTDAGAIIAGCPMLLQLWAAERFAIGRPVVDS
APYGVGRSAQWPEDGPTMGTYWCRRGRRYAHVQVRRGYPDFVFEFDRLQPSDVIWEPY
TEEAVAARAPLGLSSLCTRDQAYWLTILPMVFDIFVEPHCPQRVMRQFGLRQVFPGNV
QPTVLPADHSLTRRGQLAGALWAPRVQQYVDDWVLATEEVINELFPHTEENYRDYLRW
YLPRTRARVTFTPDAPEPHVAAVTDAYPTHRDRDYFVGADAARDISADITAVQVRLNR
GLHLTDVEQRVTFDRMQEKMRAVMRVFSCRSAVDIVPPAGPDLVCLRAPLASEQCDLQ
HRFRTTYRPMCNFIEWVDMENPQNDGTRAYPRSETRSDYLRRKDEHECRIAAEALEWH
VNPLGLPTWRERPECRCGDRCHVIRSVRHRTRGQRCFVCPNIVDDDFVFIHAWMEYER
RVEIASLDWRNNPLGLPKWSERPKCGCRDRWQVATSFAEQTFRRRYFVCPNVDNDFVV
CLTIYKIRYDLHKITTQTE"

LTR 16155. .16469
/note="3' LTR"

CDS join(23418. .23748,25558. .25589,25715. .26480,26838. .26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92399.1"
/db_xref="GI:7228439"
/translation="MGANCCIAAKERPQPCVTPIEVSAFRNVRHSPSWSFRWDNRTHI
EDIMEMPALFSNHSSGSIRPETKSGSIAPTDGFSNGGSPSDMFNKLKCHKSDRKRESS
KIARSDLRGRSTTSNSSPEAKLSRKSLDTVSVASDSKMSISVPSTPPAISRADPSSSS
RGHSLPTDADSMRKARRSPGYQLYRQVSDSKIPSLRSLNEGASPEGRPSSSSMLSVCSN
DLSAVGSHGESSDGWSMRTFSEMVASSQRERWSVDSELLGSVSSKMTRSNASNNPTTH
SPDQEVCKLCLKLLKERSTWNAQELAVVAVLLCGHVYHADCLDSLTAEADKYDPPCPV
CTHGEKCTVKLFGKLESKTKNKIPKNVIVDVNLDGSSKHQKEKIPLEEELYSQARASM
RELKSSRIGANDMTWLLNDLRIEGFS"

CDS complement(join(27953. .28624,29565. .29918,30369. .30646,
31026. .31191))
/note="ESTs C96615(C10106),C26336(C12127),D21959(C10106)
correspond to a region of the predicted gene.
Similar to NAM (AL021889)"
/codon_start=1
/protein_id="BAA92400.1"
/db_xref="GI:7228440"
/translation="MESLRDMVLPPGFGFHPKDTELISHYLKKKIHGQKIEYEIIPEV
DIYKHEPWDLPAKCDVPTQDNKWHFFAARDRKYPNGSRSNRATVAGYWKSTGKDRAIK
MGKQTIGTKKTLVFHEGRPPTGRRTEWIMHEYYIDERECQACPDMKDAYVLCRITKRN
DWIPGNGNELDNSDPHPEPYDAPPSVISTEQLNPAAEPVVGVEAAPVTVAEPDGVTTS
AITANIPSPSDDINLDDWLNELFDPFFDPEQSLASADLSPDEQNVESSNVGALAPKVE
QDYSSPNENVVDDTEYLLPEDVYNILHPGTDDFNMLQNPLDQYPIQYATDVWSGIQKE
ELWSPQANAEPSQSNEAADNGIIRRYRSMKTPETSVPQFKGKTQAKMRVGINKMATSS
SESINQTIKFENSGRLVEHQKNQAHDVASTKRSDAGKPSTELSSNRGFLRGIRNAFAG
CSDARWNMILVAGFAIGVAVVALHIGQRLGLSQRDQQHT"

CDS complement(38747. .38998)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92401.1"
/db_xref="GI:7228441"
/translation="MMGHPYLRADTARPDNWWAVPGPEVRHVDRHGTPEVPIVLCRPD
DRGPCRASLMAMRMLGLELVHFGCCYDDEPHGHPSGTSI"

LTR complement(39315. .39726)
/note="3' LTR"

CDS complement(join(40209. .40396,40917. .43545))
/note="EST C28952(C62945) corresponds to a region of the
predicted gene.
Similar to maize transposon MuDR mudrA protein isolog.
(AC003981)"
/codon_start=1
/protein_id="BAA92402.1"
/db_xref="GI:7228442"
/translation="MSSKVTFQIVHGEGNIRFGPDGVDLSDFVMTSKGIDRPAERTFQ
SIYSWLLRGFRIDQEVYTMSVSVVVSRATEGYFWELMPMDSTDAWRRYVEMAFEWSWP
LVIFVSVQEKDINVSMQTEDVEGPINAGDVVGPSMQNEENQPREEQAMGMADEGERVG
IIVDEMEREDSDNEEVDDDASSDEEGDVMATDWANEDFSGLVISEGDHVPWEYKENEV
IEGARYAHKDEMKEAVKHWAVSLQREFRVVKSTNYVYEVRCMKEDCPWRVHAYKGKWN
DYWKVSIVTEHKCYLQGVEKYHRNITSAFMASEMYSSVVGNIGFEPKSIIRHIENKFK
YTISYARAWRAKQKIIEMRYGTFEASYDNLPRLLAIIAQRNNNTYYDLHTFTSVDDRT
KSVLQRAFFSLGACINAFVHCRPVLCIDGTFMTGKYRGQILTAIGCNGNNQVLPMALA
FVESENTESWYWFLERVHIAVVRMRPNVCLIHDRHAGMLRAIDYLQNGWDEKGLPAKW
PDVRSRWCMRHMGANFYKQFKNKHLMELFKRLCAQNQEKKFNELWDKLDELTTKQTDE
QSRRPLVEGDEPPIPLGALHDDPPTMRRRSGSSIRNFTQWIENEPKEKWSLLFDTDGS
RYGIMTTNLAEVYNWVMRGVRVLPLVAIVEFILHGTQAYFRDRYKKIGPSMADDNIVF
GNVVTKYMEDKIKKARRHRVVAQGTQVHRYEIMCVDRSRRGIYRKQAVQECVLKADGG
CTCSCMKPKLHHLPCSHVLAAAGDCGIPPNVYVSNYFRKEAIFHTWSEEIYGFGISGS
YTTLSAQVFYIPDPSKLRVKKGRRQTRRIRNDMDESEAGGRTLRCSKCDLRGHTYKKC
PKNAEVPSGADASPSGQASDVLFIHMVLEQLMLTTTHFRRCTVVSDPVARTVHNSTHE
KRQLELSLSTFGAAYLHKVNHAHA"

LTR complement(43999. .44385)
/note="5' LTR"

CDS join(44572. .44723,44846. .44938,46366. .46378)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92403.1"
/db_xref="GI:7228443"
/translation="MALPRGVRSPKPMAAAAASASPPQPRPGHMLVLGTGFVGRYVSQ
RLLAQGWRVSGTCTSPAKKTELEMLGMDASIFDATSSRCVW"

CDS complement(join(48119. .48257,48429. .48703,49975. .50139))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92404.1"
/db_xref="GI:7228444"
/translation="MALVNDKDSEDDVNAPACARRAAPLVIGAPWGFVLQLARRLGEC
EPRDFRARKMAIISFCALSLLFSKNTRIYIVLLCYLRKNYRPLIVTKESKRTKCCRIT
ASLASRHGDELVCLCRARLASSALDRVVFKILKTSDFVLMVLKRCICVINPVHISRTH
GSSMYSTLPDAVPYRSTERLAGGIVIDCTIGG"

CDS complement(join(51918. .52637,52684. .53346))
/note="Similar to Arabidopsis thaliana chromosome 1 YAC
YUP8H12R sequence, C2-HC type zinc finger protein
C.e-MyT1. (AC002986)"
/codon_start=1
/protein_id="BAA92405.1"
/db_xref="GI:7228445"
/translation="MASCINGGELHVRVVHRRLVKASDESIRPHVLAVSNLDLIPRTI
QVSMFCIYPKPSTGGDFHDVVAAFAAGLPSLLNHFFPLAGRIVSNPCSGLPEIHCHNQ
GAELVVGEADVALASLDYGTVGASVGKILLPYAGDVALSVQVVSFACGGFTVAWGTNH
VVVDGSALSMLVSAWSELARSGTLAAGARPNHDRSVFRPRSPPSYGASLDEAFTPLDG
ARQAADIARLREQARATRVQAVSAYLWKALAAVVGSRDARCRMVWWVDGRRRLTLSSS
PELRAAMRSYVGNVTTFAVAEATMEEIQRKPLAEVASMARDAIAAPAYGEHFQELVDW
VEEHKAGKQRYIDTASVGLGSPAVSVTAFASFEVDTDFGHGHAAMALPTSSSSARLCT
GFVQIAARPGGGDGSWIASALLWPRLAAALESDERLGRIFKPVTAEYLGLRPASGSSA
KRAGMITSKI"

LTR complement(57570. .58009)
/note="3' LTR"

CDS complement(join(58018. .58891,58980. .62098))
/note="Similar to Arabidopsis thaliana chromosome II BAC
F9O13 genomic sequence, putative retroelement pol
polyprotein. (AC006248)"
/codon_start=1
/protein_id="BAA92406.1"
/db_xref="GI:7228446"
/translation="MRVNLQAQGLWTAIDPGYAEFREDWAALSAILQAVPREMLRGLA
KHDTAKAAWDAIKTMRVGVDRVREAKEQGFRRQFESMRFKERETPEEFAMRLTAVVAD
IRDMGGVMEDEHVNKKLLRVVPKKYKPVAISLEQLLDVKTMALEELVGRLSTVDSYSD
DEEGSDGGKLYLTEEQWQARVKQREQEGSGNSGNKGRGAPGTQNHRGKPGGSPKGKEA
ATGANSSRDISRVKCFNCDEFGHYARQCRKPRRQRRGEANLVQAAEEEPTLLMAHVVG
VSLAGEATLGRTPGGQEVHLTEKKVILDHEDGGEEEVTRDWFLDTGATNHMTGVRSAF
AELDTGVVGTVKFGDGSVIEIQGRGTVVFRCKNGDHRSLDTVYYIPKLRKNIISVGRL

alignment_scores:
Quality: 64.00 Length: 9
Ratio: 8.000 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 88.889

alignment_block:
US-09-444-281-36 x AP001366 ..

Align seg 1/1 to: AP001366 from: 1 to: 146081

3 ArgTrpProTrpTrpProTrpArgArg 11
|||||||||||||||| |||||||||

74276 CGTTGGCCATGGTGGCGTTGGAGGAGG 74302